Notes on Stochastic Population Modeling*

Jay R. Walton
Department of Mathematics
Texas A&M University
September 27, 2009

1 Notes on Probability

1.1 Basic notation and results

The following notation is adopted in this section:

i. A probability space consists of $(\Omega, \Sigma, \mathcal{P})$, where $\Omega$ is the space of outcomes, $\Sigma$ is the $\sigma$-algebra of events and $\mathcal{P}$ is a countably additive probability measure $\mathcal{P}: \Sigma \rightarrow [0, 1]$, sending $A \in \Sigma$ to $\mathcal{P}(A) \in [0, 1]$.

ii. A random variable is a measurable function $\eta: \Omega \rightarrow \mathbb{R}$.

iii. The cumulative distribution function (cdf) (of the random variable $\eta$) $F: \mathbb{R} \rightarrow [0, 1]$ is defined by $F(x) = \mathcal{P}(\{\omega | \eta(\omega) \leq x\}) = \mathcal{P}(\eta \leq x)$. In particular, $\mathcal{P}(a < \eta < b) = F(b) - F(a)$.

iv. The probability density function $f(x) := F'(x)$ is the derivative of the distribution function, and hence $F(b) - F(a) = \int_{a}^{b} f(x) dx$. Examples:

1. The uniform distribution on $[0, 1]$ is given by $F(x) = x$, hence $f(x) = 1$ for $0 \leq x \leq 1$.

2. The normal distribution $f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-(x-\mu)^2/(2\sigma^2)}$.

3. The exponential distribution $f_{\alpha}(x) = \frac{1}{\alpha} e^{-x/\alpha}$.

vi. Expected Value: If $\eta$ is a random variable with distribution function $F(x)$ and $g(x)$ is a real valued function, then the Expected Value, $E[g(\eta)]$, of the random variable $g(\eta)$ is defined to be

$$E[g(\eta)] := \int_{-\infty}^{\infty} g(x) dF(x).$$

v. Moments: The $n^{th}$ moment of a distribution $F$ is defined as $E[x^n] = \int_{-\infty}^{\infty} x^n dF$.

---

*Copyright © 2008 by J. R. Walton. All right reserved.
1. The first moment of the distribution $F$ is called the \textit{mean} and defined by

$$
\mu := E[x] = \int_{-\infty}^{\infty} x \, dF.
$$

For the distributions described above one has:

(a) The uniform distribution has $\mu = \int_{0}^{1} x \, dx = 1/2$.

(b) The normal distribution $f_{\mu,\sigma}$ has mean $\mu$, i.e. $\int_{-\infty}^{\infty} x f_{\mu,\sigma}(x) \, dx$.

(c) The exponential distribution $f_{\alpha}$ has mean $\mu = \frac{1}{\alpha} \int_{0}^{\infty} xe^{-x/\alpha} \, dx = \frac{\alpha}{1}$.

vi. The 2\textsuperscript{nd} \textit{moment about the mean}, called the \textit{variance}, is defined as:

$$
\text{Var}[\eta] = \sigma^2 := E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \, dF(x).
$$

In the examples above:

(a) $\sigma^2 = \int_{0}^{1} (x - 1/2)^2 \, dx = 1/12$.

(b) $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/(2\sigma^2)} \, dx$

(c) $\sigma^2 = \int_{0}^{\infty} (x - \alpha)^2 e^{-x/\alpha} \, dx = \alpha^2 \int_{0}^{\infty} (t - 1)^2 e^{-t} \, dt = \alpha^2$

vii. \textit{Poisson distribution}: Discrete exponential $\eta: \Omega \rightarrow \{0, 1, 2, \ldots\}$ (describing the number of events occurring per unit of time or per unit of space). Here $\mathcal{P}(x; \lambda) = \mathcal{P}[\eta = x] = e^{-\lambda} \frac{\lambda^x}{x!}$, and hence $\mu = E[\eta] = \lambda$ and $\sigma^2 = \text{Var}[\eta] = \lambda$.

Additional terminology:

A. Independent events: $\mathcal{P}(A \cap B) = \mathcal{P}(A) \mathcal{P}(B)$.

B. Independent random variables: $\mathcal{P}(\eta_1 \leq x \text{ and } \eta_2 \leq y) = \mathcal{P}(\eta_1 \leq x) \mathcal{P}(\eta_2 \leq y)$

C. Joint distribution of two random variables: $F_{\eta_1 \eta_2}(x, y) := \mathcal{P}(\eta_1 \leq x \text{ and } \eta_2 \leq y)$

D. Joint density: $f_{\eta_1 \eta_2}(x, y) = \partial_{xy}^2 F_{\eta_1 \eta_2}(x, y)$. Therefore

$$
\mathcal{P}(\eta_1 \leq x \text{ and } \eta_2 \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{\eta_1 \eta_2}(s, t) \, ds \, dt = F_{\eta_1 \eta_2}(x, y)
$$

E. If $\eta_1$ and $\eta_2$ are independent random variables, then $F_{\eta_1 \eta_2}(x, y) = F_{\eta_1}(x)F_{\eta_2}(y)$ and $f_{\eta_1 \eta_2}(x, y) = f_{\eta_1}(x)f_{\eta_2}(y)$.

F. Covariance: $\text{Cov}(\eta_1, \eta_2) := E[(\eta_1 - E(\eta_1))(\eta_2 - E(\eta_2))] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_1)(y-\mu_2) F_{\eta_1 \eta_2}(x, y) \, dx \, dy$.

Therefore, if $\eta_1$ and $\eta_2$ are independent random variables, then $\text{Cov}(\eta_1, \eta_2) = 0$. The reverse implication is in general not true. However, as an important special case, two uncorrelated Gaussian random variables are also independent.

G. If $f_{\eta_1 \eta_2}(x, y)$ is the joint probability density function, then $f_{\eta_1}(x) = \int_{-\infty}^{\infty} f_{\eta_1 \eta_2}(x, s) \, ds$ and $f_{\eta_2}(y) = \int_{-\infty}^{\infty} f_{\eta_1 \eta_2}(t, y) \, dt$
Conditional probability. Given random variables \( \eta_1 \) and \( \eta_2 \) and constants \( \lambda_1 \) and \( \lambda_2 \), one has

\[
E[\lambda_1 \eta_1 + \lambda_2 \eta_2] = \lambda_1 E[\eta_1] + \lambda_2 E[\eta_2].
\]

Lemma 1.1. The expected value is a linear function of the random variables. In other words, given random variables \( \eta_1 \) and \( \eta_2 \) and constants \( \lambda_1 \) and \( \lambda_2 \), one has

\[
E[\lambda_1 \eta_1 + \lambda_2 \eta_2] = \lambda_1 E[\eta_1] + \lambda_2 E[\eta_2].
\]

Lemma 1.2. If \( \eta_1 \) and \( \eta_2 \) are independent random variables, then

\[
\text{Var}[\lambda_1 \eta_1 + \lambda_2 \eta_2] = |\lambda_1|^2 \text{Var}[\eta_1] + |\lambda_2|^2 \text{Var}[\eta_2].
\]

Lemma 1.3 (Chebyshev inequality). Suppose \( g(x) \geq 0 \) is a non-decreasing function, i.e. \( a < b \) implies \( g(a) \leq g(b) \). Then \( \mathcal{P}(\eta \geq a) \leq E[g(\eta)]/g(a) \).

**Proof.** \( E[g(\eta)] = \int_{-\infty}^{\infty} g(x) dF \geq \int_{a}^{\infty} g(x) dF \geq g(a) \int_{a}^{\infty} dF = g(a) \mathcal{P}(\eta \geq a) \).

Example 1. Define \( g(x) = \begin{cases} x^2 & x \geq 0 \\ 0 & x < 0 \end{cases} \) and denote \( \xi := |\eta - E[\eta]| \). Then \( \mathcal{P}(\xi \geq a) \leq \frac{E[g(\xi)]}{g(a)} \). Since \( E[g(\xi)] = \text{Var}(\eta) \), one gets

\[
\mathcal{P}(\xi \geq a) \leq \frac{\text{Var}(\eta)}{a^2}
\]

Let \( a = \sigma k \), then \( \mathcal{P}(\xi \geq \sigma k) \leq \frac{\text{Var}(\eta)}{\sigma^2 k^2} = 1/k^2 \).

Next is the celebrated Central Limit Theorem.

**Theorem 1.1.** Let \( \eta_1, \eta_2, \ldots \) be a sequence of independent identically distributed random variables with mean 0 and variance 1 and pdf(\( \eta_j \)) = \( f \), \( j = 0, \ldots, n \). Define \( S_n := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j \) and denote \( f_{S_n} := \text{pdf}(S_n) \). Then

\[
\lim_{n \to \infty} \int_{-\infty}^{b} f_{S_n}(x) dx = \int_{-\infty}^{b} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx
\]

**QUESTION:** If \( \eta \) and \( \xi \) are independent random variables with densities \( f_\eta \) and \( f_\xi \), respectively, what is the density of \( \eta + \xi \)?

\[
\mathcal{P}(\eta + \xi \leq x) = \int \int_{s+t\leq x} f_\eta(s)f_\xi(t) ds dt = \int_{-\infty}^{\infty} ds \int_{-\infty}^{x-s} f_\eta(s)f_\xi(t) dt \\
= \int_{-\infty}^{\infty} ds \int_{-\infty}^{x} f_\eta(s)f_\xi(r-s) dr = \int_{-\infty}^{x} dr \int_{-\infty}^{\infty} f_\eta(s)f_\xi(r-s) ds.
\]

Hence \( f_{\eta+\xi}(x) = \int_{-\infty}^{\infty} f_\eta(x) f_\xi(x-s) ds = f_\eta \ast f_\xi (x) \).

**Conditional probability.** If \( A \) and \( B \) are events, define the conditional probability of \( B \) given \( A \) \( \mathcal{P}[B|A] := \mathcal{P}[A \cap B]/\mathcal{P}[A] \). Note: If \( A \) and \( B \) are independent, the \( \mathcal{P}[B|A] = \mathcal{P}[B] \).
Conditional expected value. There are multiple routes to defining the notion of the “expected value of one RV given a second RV”. One such route begins with defining the “expected value of one RV given an event”. If $\eta$ is a random variable, then

$$E[\eta|A] := \int_{-\infty}^{\infty} \eta(\omega) \mathcal{P}(d\omega|A).$$

Next, let $Z = \{Z_j\}_{j=1}^\infty$ be a (measurable) partition of $\Omega$, i.e. $\Omega = \bigcup_{j=1}^\infty Z_j$ and $Z_i \cap Z_j = \emptyset$ if $i \neq j$.

**Definition of $E[\eta|Z]$**. Given an RV $\eta$, define the random variable

$$E[\eta|Z](\omega) := \sum_{j=1}^\infty \sum_{j=1}^\infty \chi_{Z_j}(\omega) E[\eta|Z_j].$$

Then one defines $E[\eta|\xi]$ by

**Definition of $E[\eta|\xi]$**. Let $\eta$ and $\xi$ be random variables. Assume first that $Z = \{Z_1, \ldots, Z_n\}$ is a partition of $\Omega$ and $\xi = \sum_{j=1}^n \xi_j \chi_{Z_j}(\omega)$, i.e. the RV $\xi$ takes only a finite number of values. Define $E[\eta|\xi] := E[\eta|Z]$. Note that $E[\eta|\xi]$ is a random variable obtained as a function of the random variable $\xi$.

**Theorem 1.2**. Let $h(\xi)$ be any measurable function of $\xi$. Then

$$E[(\eta - E[\eta|\xi])^2] \leq E[(\eta - h(\xi))^2].$$

**Proof**. Recall that $\int_0^1 (f(x) - c)^2 dx$ is minimized for $c = \int_0^1 f(x) dx$. Therefore,

$$E[(\eta - h(\xi))^2] = \int_\Omega (\eta - h(\xi))^2 \mathcal{P}(d\omega) = \sum_i \mathcal{P}(Z_i) \int_{Z_i} (\eta - h(\xi))^2 \frac{\mathcal{P}(d\omega)}{\mathcal{P}(Z_i)}.$$

Note that for each $i$, $\int_{Z_i} (\eta - h(\xi))^2 \frac{\mathcal{P}(d\omega)}{\mathcal{P}(Z_i)}$ is minimized for $h(\xi_i) = E[\eta|Z_i]$.

Define

$$\mathcal{L}^2[(\Omega, \Sigma, \mathcal{P})] = \{\eta \mid E[\eta^2] = \int_\Omega \eta^2(\omega) \mathcal{P}(d\omega) < \infty\},$$

and define the $\mathcal{L}^2$ inner product

$$\langle \eta_1, \eta_2 \rangle := E[\eta_1 \eta_2],$$

for $\eta_1, \eta_2 \in \mathcal{L}^2[(\Omega, \Sigma, \mathcal{P})]$.

Let

$$\mathcal{S}_\xi := \{\eta \in \mathcal{L}^2[(\Omega, \Sigma, \mathcal{P})] \mid \eta = h(\xi(\omega))\}$$

denote the linear subspace consisting of those random variables in $\mathcal{L}^2$ which are functions of $\xi$. Theorem 1.2 shows that $E[\eta|\xi]$ is the image of $\eta$ under the orthogonal projection operator from $\mathcal{L}^2[(\Omega, \Sigma, \mathcal{P})]$ to $\mathcal{S}_\xi$, denoted by $\mathbb{P}_\xi$. Then one can take as the definition of $E[\eta|\xi]$

$$E[\eta|\xi] := \mathbb{P}_\xi \eta.$$
Streamlined approach

Alternatively, one can define $E[\eta|\xi]$ directly in terms of the joint distribution of the two RVs $\eta$ and $\xi$ as follows.

**Definition 1.**

$$E[g(\eta, \xi)|\xi] := \frac{\int_{-\infty}^{\infty} g(s, \xi) f_{\eta \xi}(s, \xi)ds}{\int_{-\infty}^{\infty} f_{\eta \xi}(s, \xi)ds} = \frac{\int_{-\infty}^{\infty} g(s, \xi) f_{\eta \xi}(s, \xi)ds}{f_\xi(\xi)}$$

where $f_{\eta \xi}$ is the joint pdf of $\eta$ and $\xi$.

Note that $f_\xi(t) = \int_{-\infty}^{\infty} f_{\eta \xi}(s, t)ds$.

**Special case** $g(\eta, \xi) = g(\eta)$: In this case

$$E[g(\eta)|\xi] = \frac{\int_{-\infty}^{\infty} g(s) f_{\eta \xi}(s, \xi)ds}{f_\xi(\xi)}.$$ 

In particular,

$$E[\eta|\xi] = \frac{\int_{-\infty}^{\infty} s f_{\eta \xi}(s, \xi)ds}{f_\xi(\xi)}.$$ 

**Conditional distribution:** Let $\eta, \xi$ be random variables. Define

$$f_{\eta|\xi=y}(x) := \frac{f_{\eta \xi}(x, y)}{f_\xi(y)} \quad \text{and} \quad f_{\eta=x|\xi}(y) := \frac{f_{\eta \xi}(x, y)}{f_\eta(x)}$$

It follows that $P(\eta \in A | \xi = y) = \int_A f_{\eta|\xi=y}(x)dx$, and hence

$$E[\eta | \xi = y] = \mu_{\eta|\xi=y} = \int_{-\infty}^{\infty} x f_{\eta|\xi=y}(x)dx.$$ 

**Bayes' Theorem:**

$$P[A|B] = \frac{P[B|A]P(A)}{P(B)},$$ 


Now, let $\mathcal{Z} = \{Z_j\}$ be a countable partition of $\Omega$. Then $P[A] = \sum_j P[A \cap Z_j] = \sum_j P[A|Z_j]P[Z_j] = \sum_j P[Z_j|A]P[A]$. Therefore:

$$P[Z_j|A] = \frac{P[A|Z_j]P[Z_j]}{\sum_k P[A|Z_k]P[Z_k]}.$$
1.2 Stochastic processes

Let \( \eta \) be a random variable, \( \{ \eta(t) | t \in T \subset \mathbb{R}^+ \} \).

**Counting Stochastic Processes. Axioms:**

1. \( \eta(t) \in \{0, 1, 2, \ldots \} = \mathbb{Z}^+ \cup \{0\} \).
2. If \( s < t \) then \( \eta(t) - \eta(s) \geq 0 \).
3. If \( s < t \) then \( \eta(t) - \eta(s) \) = number of events occurring in \( (s, t] \).

**SPECIAL CASE: Poisson processes**

4. \( \eta(0) = 0 \)
5. \( \eta(s) \) and \( \eta(t+s) - \eta(s) \) are independent random variables for all \( s, t \geq 0 \).
6. (Stationary increments) Distributions of \( \eta(t+s) - \eta(s) \) depends only upon \( t \).
7. \( \mathcal{P}(\eta(t+h) - \eta(t) = 1) = \lambda h + o(h) \) as \( h \to 0^+ \) and \( \mathcal{P}(\eta(t+h) - \eta(t) > 1) = o(h) \) as \( h \to 0^+ \).

**Proposition 1.1.** In a Poisson process one has \( p(x; t) := \mathcal{P}(\eta(t) = x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!} \) (Poisson distribution).

**Proof.**

\[
p(0; t + h) = \mathcal{P}(\eta(t + h) = 0) = \mathcal{P}(\eta(t + h) - \eta(t) = 0 \& \eta(t) = 0)
= \mathcal{P}(\eta(t + h) - \eta(t) = 0 \& \eta(t) = 0)
= \mathcal{P}(\eta(t + h) = \eta(t) = 0) \mathcal{P}(\eta(t) = 0) = \mathcal{P}(\eta(h) - \eta(0) = 0) \mathcal{P}(\eta(t) = 0)
= \mathcal{P}(\eta(h) = 0) \mathcal{P}(\eta(t) = 0) = p(0; h)p(0; t)
= p(0; t)(1 - \mathcal{P}(\eta(h) > 0))
= p(0; t)(1 - \lambda h) + o(h).
\]

Sending \( h \to 0 \) one gets \( \partial_t p(0; t) = -\lambda p(0; t) \), hence \( p(0; t) = e^{-\lambda t} \).

For \( x \geq 1 \),

\[
\mathcal{P}(\eta(t + h) = x) = p(x; t + h) = \mathcal{P}(\eta(t) = x \& \eta(t + h) - \eta(t) = 0) + \mathcal{P}(\eta(t) = x - 1 \& \eta(t + h) - \eta(t) = 1)
+ \sum_{k=2}^{x} \mathcal{P}(\eta(t) = x \& \eta(t + h) - \eta(t) = k)
= \mathcal{P}(\eta(t) = x) \mathcal{P}(\eta(t + h) - \eta(t) = 0) + \mathcal{P}(\eta(t) = x - 1) \mathcal{P}(\eta(t + h) - \eta(t) = 1)
+ \sum_{k=2}^{x} \mathcal{P}(\eta(t) = x - k) \mathcal{P}(\eta(t + h) - \eta(t) = k)
= p(x; t)e^{-\lambda h} + p(x - 1, t)\lambda h + 0(h).
\]
It follows that
\[
\lim_{h \to 0} \frac{p(x; t + h) - p(x; t)e^{-\lambda h}}{h} = \lambda p(x - 1; t) = \partial_x p(x; t) + \lambda p(x; t). \tag{1}
\]

Therefore, \( \frac{d}{dt} (e^{\lambda t} p(x; t)) = \lambda e^{\lambda t} p(x - 1; t) \) and hence
\[
\int_0^t \frac{d}{dr} (e^{\lambda r} p(x; r)) dr = \int_0^t \lambda e^{\lambda r} p(x - 1; r) dr = e^{\lambda t} p(x; t) - p(x; 0).
\]

One concludes that \( p(x; t) = e^{-\lambda t} p(x; 0) + e^{-\lambda t} \int_0^t \lambda e^{\lambda r} p(x - 1; r) dr. \) Now, apply induction. When \( x = 1, \)
\[
p(1; t) = e^{-\lambda t} p(1; 0) + e^{-\lambda t} \int_0^t \lambda e^{\lambda r} p(0; r) dr = \lambda e^{-\lambda t} \int_0^t \lambda e^{\lambda r} e^{-\lambda r} dr = \lambda e^{-\lambda t}.
\]

Assume the result for some \( x \geq 1. \) Then
\[
p(x + 1; t) = e^{-\lambda t} p(x + 1; 0) + e^{-\lambda t} \int_0^t \lambda e^{\lambda r} p(x; r) dr
\]
\[
= \lambda e^{-\lambda t} \int_0^t e^{\lambda r} e^{-\lambda r} \frac{(\lambda r)^x}{x!} dr
\]
\[
= \lambda x! e^{-\lambda t} \int_0^t \frac{r^x}{x!} dr
\]
\[
= \lambda x! e^{-\lambda t} \frac{x^{x+1}}{x!}.
\]

*******************************

Alternative argument:

It follows from (1) that, for \( x \geq 1, \) \( p(x; t) \) is a solution to the following initial value problem:
\[
\dot{\varphi} + \lambda \varphi = p(x - 1; t), \quad \varphi(0) = 0.
\]
One can proceed using induction. We have seen that \( p(0; t) = e^{-\lambda t}. \) Assuming that \( p(x - 1; t) = e^{-\lambda t} \frac{(\lambda)^{x-1}}{x!}, \) then \( p(x; t) = e^{-\lambda t} \frac{(\lambda)^x}{x!} \) is the only solution to the initial value problem above. This concludes the proof.

2 Stochastic population model

In the deterministic model the population growth is governed by the differential equation
\[
\dot{u}(t) = [B(u(t)) - D(u(t))] u(t),
\]
where \( B(u) \) and \( D(u) \) are birth and death rates, respectively, expressed as functions of the current population size,
In the **stochastic population model**, one models the population at time \( t \) as a stochastic process \( \eta(t) \) and defines \( \eta(x; t) = P(\eta(t) = x) \), with \( x \in \{0, 1, 2, \ldots \} \) and initial condition \( \eta(0) = u_0 \). Thus, rather than predicting that the population at time \( t \) has the specific value \( u(t) \), as in the deterministic model, the stochastic model only predicts the probability that the population has the specific value \( x \) at time \( t \).

The basic assumption connecting the deterministic and stochastic population models is that \( \eta(t) \) is a continuous (in time) counting process satisfying the following conditional probability axioms:

1. \( P(\eta(t + h) = x + 1|\eta(t) = x) = xB(x)h + o(h) \)
2. \( P(\eta(t + h) = x - 1|\eta(t) = x) = xD(x)h + o(h) \)
3. \( P(\eta(t + h) = x|\eta(t) = x) = (1 - x(B(x) + D(x)))h + o(h) \)

where \( o(h) \) (pronounced “little oh \( h \)”) is the Landau order convention whereby the symbol \( o(h) \) denotes a function satisfying the asymptotic condition

\[
\lim_{h \to 0} \frac{o(h)}{h} = 0.
\]

Under these assumptions, one derives a set of conditions that must be satisfied by \( p(x; t) \).

First observe that

\[
p(x; t + h) := P[\eta(t + h) = x]
= P[\eta(t + h) = x|\eta(t) = x + 1] \cdot P(\eta(t) = x + 1) + P[\eta(t + h) = x|\eta(t) = x] \cdot P(\eta(t) = x)
= P[\eta(t + h) = x|\eta(t) = x - 1] \cdot P(\eta(t) = x - 1)
= (x + 1)D(x + 1)hp(x + 1; t) + (1 - (B(x) + D(x))xh)p(x; t) + B(x - 1)hp(x - 1; t) + o(h).
\]

As \( h \to 0^+ \), one obtains

\[
\partial_t p(x; t) = (x + 1)D(x + 1)p(x + 1; t) - (B(x) + D(x))xp(x; t) + (x - 1)B(x - 1)p(x - 1; t).
\]

This infinite dimensional set of ordinary differential equations is called the **Kolmogorov Equations**.

In particular, for \( x = 0 \) one obtains

\[
\partial_t p(0; t) = D(1)p(1; t).
\]

**Definition 2.** The **Probability Generating Function** for this process is defined as

\[
P(s; t) := \sum_{x=0}^{\infty} s^x p(x; t),
\]

hence, \( p(x; t) = \frac{1}{x!} \partial_s^x P(s; t)|_{s=0} \).

Thus, in principle one can find the probabilities \( p(x; t) \) if one knows the generating function \( P(s; t) \). However, in practice one must be able to take partial derivatives of \( P(s; t) \) of arbitrary order to find all of the probabilities \( p(x; t) \).
Example 2. In a Poisson process, the probability generating function is \( P(s; t) = e^{\lambda t (s-1)} \).

One strategy for finding the generating function \( P(s; t) \) is to derive a partial differential equation that it satisfies. To that end, suppose one writes

\[
\partial_t p(x; t) = A_{-1}(x-1)p(x-1; t) + A_0(x)p(x; t) + A_1(x+1)p(x+1; t).
\]

Then

\[
\partial_t P(s; t) = \sum_{x=0}^{\infty} s^x \partial_t p(x; t) = \sum_{x=1}^{\infty} s^x \sum_{j=-1}^{1} A_j(x+j)p(x+j; t),
\]

where \( p(-1; t) = 0 \).

On the other hand, \( \partial_s P(s; t) = \sum_{x=1}^{\infty} x s^{x-1} p(x; t) \), and hence \( \partial_s P(s; t) = \sum_{x=0}^{\infty} x s^x P(x; t) \).

It follows that

\[
(s \partial_s)^k P(s; t) = \sum_{x=0}^{\infty} x^k s^x P(x; t).
\]

To see how this can lead to a partial differential equation for \( P(s; t) \) consider the following example.

Suppose \( A_j(x+j) = (x+j)^K \), some \( K \geq 1 \) and adopt the convention that \( p(-1; t) = 0 \).

**Case 1.**

\[
\sum_{x=0}^{\infty} s^x (x-1)^K p(x-1; t) = s \sum_{x=0}^{\infty} s^x x^K p(x; t) = s (s \partial_s)^K P(s; t).
\]

**Case 2.**

\[
\sum_{x=0}^{\infty} s^x x^K p(x; t) = (s \partial_s)^K P(s; t).
\]

**Case 3.**

\[
\sum_{x=0}^{\infty} s^x (x+1)^K p(x+1; t) = \frac{1}{s} \sum_{x=0}^{\infty} s^{x+1} (x+1)^K p(x+1; t) = \frac{1}{s} \left[ \sum_{x=0}^{\infty} s^x x^K p(x; t) \right] = \frac{1}{s} (s \partial_s)^K P(s; t).
\]

It follows from (5) that under these assumptions one has

\[
\partial_t P(s; t) = \left( s + 1 + \frac{1}{s} \right) (s \partial_s)^K P(s; t).
\]

This is the desired partial differential equation for \( P(s; t) \). Unfortunately, in general there is no simple method for constructing its solution.

It is instructive to study the following important example stochastic population model.

**Example 3. (Logistic Model)** Consider the stochastic population model corresponding to the classical logistic deterministic model \( \dot{u} = ru(1-u/k) \), i.e. \( B(u) = r \) and \( D(u) = \frac{r}{k} u \).

Then the resulting Kolmogorov Equations are

\[
\partial_t p(x; t) = (x+1)^2 \frac{r}{k} p(x+1; t) - r(1+x/k) x p(x; t) + (x-1) r p(x-1; t).
\]

Hence \( A_{-1}(x-1) = r(x-1) \), \( A_0(x) = rx(1+x/k) \) and \( A_1(x+1) = r/k (x+1)^2 \).

Therefore
1. $\sum_{x=0}^{\infty} s^x A_{-1}(x-1)p(x-1; t) = r \sum_{x=0}^{\infty} s^x (x-1)p(x-1; t) = rs(s \partial_s)P(s; t)$;

2. $\sum_{x=0}^{\infty} s^x A_0(x)P(x; t) = r \sum_{x=0}^{\infty} s^x xp(x; t) + r/k \sum_{x=0}^{\infty} s^x x^2 p(x; t) = r(s \partial_s)P(s; t) + r/k (s \partial_s)^2 P(s; t)$;

3. $\sum_{x=0}^{\infty} s^x A_1(x+1)p(x+1; t) = r/k \sum_{x=0}^{\infty} s^x (x+1)^2 p(x+1; t) = r/(ks(s \partial_s)^2 P(s; t)$.

So, $P(s; t)$ satisfies $\partial_t P(s; t) = r(s^2 + s) \partial_s P(s; t) + r/k \ (1 + 1/s) (s \partial_s)^2 P(s; t)$, and since $(s \partial_s)^2 = s(\partial_s + s \partial_s^2)$ one gets

$$\partial_t P(s; t) = r \ (s + 1) \left( s + \frac{1}{k} \right) \partial_s P(s; t) + \frac{r}{k} \ (s + 1) \ s \partial_s^2 P(s; t).$$

This partial differential equation for $P(s; t)$ is a nonhomogeneous, parabolic equation.

While much is known about the theoretical properties of this equation, there is no simple procedure for constructing its solution in closed form.

Often it is more convenient to get information about the stochastic population process $p(x; t)$ from other generating functions such as the

**Moment Generating Function.**

Define

$$M(\theta, t) := \sum_{x=0}^{\infty} e^{\theta x} p(x; t) = \sum_{x=0}^{\infty} p(x; t) \left( \sum_{k=0}^{\infty} \frac{\theta^k x^k}{k!} \right) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} \left( \sum_{x=0}^{\infty} x^k p(x; t) \right) = \sum_{k=0}^{\infty} \mu_k(t) \frac{\theta^k}{k!}. \tag{9}$$

Here $\mu_k(t) := \sum_{x=0}^{\infty} x^k p(x; t)$ is the $k$-th moment of the pdf $p(x; t)$.

**Example 4. (Logistic Model.)** Assume that $p(x; t)$ satisfies (4). Then:

$$\partial_t M(\theta, t) = \sum_{x=0}^{\infty} e^{\theta x} \partial_t p(x; t) = \sum_{x=0}^{\infty} e^{\theta x} \left[ \sum_{j=-1}^{1} A_j(x + j)p(x + j; t) \right].$$

Note:

$$\partial_\theta^k (M(\theta, t)) = \sum_{x=0}^{\infty} e^{\theta x} x^k p(x; t)$$

$$\partial_\theta^k (e^{-\theta} M(\theta, t)) = \sum_{x=0}^{\infty} e^{\theta (x-1)} (x - 1)^k p(x; t)$$

$$\partial_\theta^k (e^\theta M(\theta, t)) = \sum_{x=0}^{\infty} e^{\theta (x+1)} (x + 1)^k p(x; t)$$

**Case i.**

$$\sum_{x=0}^{\infty} e^{\theta x} (x - 1)^k p(x - 1; t) = \sum_{x=0}^{\infty} e^{\theta (x+1)} x^k p(x; t) = e^\theta \partial_\theta^k M(\theta, t).$$

**Case ii.**

$$\sum_{x=0}^{\infty} e^{\theta x} x^k p(x; t) = \partial_\theta^k M(\theta, t)$$
Case iii.

$$\sum_{x=0}^{\infty} e^{\theta x}(x+1)k p(x+1; t) = e^{-\theta} \sum_{x=1}^{\infty} e^{\theta x} x^k p(x; t) = e^{-\theta} \sum_{x=0}^{\infty} e^{\theta x} x^k p(x; t) = e^{-\theta} \partial_{\theta}^k M(\theta, t)$$

In the logistic model, one has $A_{-1}(x-1) = r(x-1), A_0(x) = x(1 + x/k), A_1(x+1) = r/k(x+1)^2$. Therefore,

$$\partial_t M(\theta, t) = r \sum_{x=0}^{\infty} e^{\theta x}(x-1)p(x-1; t) + r \sum_{x=0}^{\infty} e^{\theta x}(x)p(x; t)$$

$$+ \frac{r}{k} \sum_{x=0}^{\infty} e^{\theta x}x^2 p(x; t) + \frac{r}{k} \sum_{x=0}^{\infty} e^{\theta x}(x+1)^2 p(x+1; t)$$

$$= re^\theta \partial_\theta M(\theta, t) + r \partial_\theta M(\theta, t) + \frac{r}{k} \partial_\theta^2 M(\theta, t) + \frac{r}{k} e^{-\theta} \partial_\theta^2 M(\theta, t)$$

$$= r(e^\theta + 1) \partial_\theta M(\theta, t) + \frac{r}{k} (e^{-\theta} + 1) \partial_\theta^2 M(\theta, t).$$

In general, $M(\theta, t) := \sum_{x=0}^{\infty} e^{\theta x} p(x; t) = \sum_{k=0}^{\infty} \mu_k(t) \theta^k$. Hence,

$$\partial_\theta M(\theta, t) = \sum_{k=0}^{\infty} \mu_k(t) k \theta^{k-1} = \sum_{k=0}^{\infty} \mu_{k+1}(t) (k+1) \theta^k, \quad (10)$$

and

$$\partial_\theta^2 M(\theta, t) = \sum_{k=0}^{\infty} \mu_k(t) k(k-1) \theta^{k-2} = \sum_{k=2}^{\infty} \mu_k(t) k(k-1) \theta^{k-2} = \sum_{k=0}^{\infty} \mu_{k+2}(t) (k+2)(k+1) \theta^k. \quad (11)$$

Also,

$$\partial_t M(\theta, t) = \sum_{k=0}^{\infty} \dot{\mu}_k(t) \theta^k. \quad (12)$$

To be well-posed, a partial differential equation must be augmented with boundary or initial conditions. For the PDE satisfied by the moment generating function, one can deduce the appropriate initial condition from the initial condition for the population process $p(x; t)$ as follows. If $\eta(0) = u_0$, then

$$p(x; 0) = \begin{cases} 1 & , x = u_0 \\ 0 & , x \neq u_0. \end{cases} \quad (13)$$

Under this assumption one has $\mu_1(0) = \sum_{x=0}^{\infty} xp(x; 0) = u_0$ and $\mu_k(0) = \sum_{x=0}^{\infty} x^k p(x; 0) = u_0^k$, and $\text{Var}(0) = \sum_{x=0}^{\infty} (x - \mu_1(0))^2 p(x; 0) = 0$.

Recall: $\text{Var}(t) = \sum_{x=0}^{\infty} (x - \mu_1(t))^2 p(x; t) = \sum_{x=0}^{\infty} (x^2 - 2\mu_1(t)x + \mu_1(t)^2) p(x; t) = \mu_2(t) - \mu_1(t)^2.$
Transform PDE for mgf to infinite ODE system for momenta $\mu_k(t), k = 1, 2, \ldots$

$$(e^\theta + 1) \partial_\theta M(\theta, t) = \left( 2 + \sum_{j=1}^{\infty} \frac{\theta^j}{j!} \right) \left( \sum_{k=0}^{\infty} \mu_{k+1}(t)(k+1)\theta^k \right) = 2\mu_1 + \theta(4\mu_2(t) + \mu_1(t)) + \theta^2 + \ldots$$

Recall: (product formula for infinite series)

Then:

$$(e^{-\theta} + 1) \partial^2_\theta M(\theta, t) = \left( 2 + \sum_{k=1}^{\infty} \frac{(-\theta)^k}{k!} \right) \left( \sum_{k=0}^{\infty} \mu_{k+2}(t)(k+2)(k+1)\theta^k \right).$$

Cumulant generating function $\leftrightarrow$ cfg

$$K(\theta, t) := \log(M(\theta, t)) = \sum_{i=0}^{\infty} K_i(t) \frac{\theta^i}{i!},$$

where $K_0(t) = 0$, $K_1(t) = \partial_\theta K(\theta, t)|_{\theta=0} = \frac{\partial_\theta M(\theta, t)}{M(\theta, t)}|_{\theta=0} = \mu_1(t)$.

and

$$K_2(t) = \partial^2_\theta K(\theta, t)|_{\theta=0} = \partial_\theta \left( \frac{\partial_\theta M(\theta, t)}{M(\theta, t)} \right)|_{\theta=0} = \mu_1(t) = \left( \frac{\partial^2_\theta M}{M^2} - \frac{(\partial_\theta M)^2}{M^2} \right)|_{\theta=0} = \mu_2(t) - \mu_1(t)^2 = E((\eta(t) - \mu_1(t))^2).$$

In general:

$$K_j(t) = E\left((\eta(t) - \mu_1(t))^j\right). \quad (14)$$

Multiple population stochastic model

The single population stochastic model discussed above can be readily generalized to the multiple interacting population setting. The language of vector valued functions is convenient for this purpose.

**Deterministic model:** For deterministic multiple population models with $n$ interacting species, one can introduce a vector valued function $u(t) = (u_i(t)), i = 1, \ldots, n$, whose components $u_i(t)$ give the population size of the $i^{th}$ species being modeled in the ecosystem. The dynamic interactions of the $n$ species are then modeled through the $n \times n$ system of differential equations

$$\dot{u}(t) = f(u(t)) = B(u(t)) - D(u(t))$$

where the components of $B(u(t))$ and $D(u(t))$ are all nonnegative representing interactions leading to growth or decline of each of the populations, respectively.

**Stochastic model:** For the stochastic counterpart to this $n \times n$ deterministic population model, one introduces a vector valued stochastic process $\eta$ satisfying the following axioms:
\begin{align*}
P(\eta_j(t + h) = x + 1 | \eta_j(t) = x) &= B_j h + o(h) \quad (15) \\
P(\eta_j(t + h) = x - 1 | \eta_j(t) = x) &= D_j h + o(h) \quad (16) \\
P(\eta_j(t + h) = x | \eta_j(t) = x) &= (1 - (B_j + D_j)) h + o(h) \quad (17)
\end{align*}

for \( j = 1, \ldots, n \). Thus, as before the rates of change from the deterministic model go into conditional probability of change in the stochastic model. One can now develop the Kolmogorov equations for this multiple population stochastic model as well as the moment generating function machinery. But it is considerably more complicated than in the single population case. However, performing numerical simulations for the stochastic multiple population model is not really more complicated than for a stochastic single population model. From Monte Carlo simulations one can learn much about the distribution properties of the vector stochastic process \( \eta(t) \) such as how the mean and variance evolve as functions of time.

2.1 Simulation of Stochastic Model

To illustrate how to perform numerical simulations of a stochastic multiple population model, consider a two population model

\[ \eta(t) = \begin{pmatrix} \eta_1(t) \\ \eta_2(t) \end{pmatrix}. \]

One must simulate both what events occur in succession and the time interval between successive events, both of which are random variables. To simulate a succession of events for the process one implements the following algorithm:

0. Specify initial time, say \( t_0 = 0 \), and population sizes:

\[ \eta(t_0) = \eta_0 = \begin{pmatrix} \eta_{10} \\ \eta_{20} \end{pmatrix} \]

i. Generate a (pseudo-)random number \( 0 \leq Y \leq 1 \);

ii. Compute \( B_1, D_1, R \) where \( R \) is defined as \( R := B_1 + D_1 + B_2 + D_2 \);

iii. If \( 0 \leq Y \leq B_1/R \), then \( \eta_1 \mapsto \eta_1 + 1 \);

iv. If \( B_1/R < Y \leq (B_1 + D_1)/R \), then \( \eta_1 \mapsto \eta_1 - 1 \);

v. If \( (B_1 + D_1)/R < Y \leq (B_1 + D_1 + D_2)/R \), then \( \eta_2 \mapsto \eta_2 + 1 \);

vi. otherwise \( \eta_2 \mapsto \eta_2 - 1 \).

vii. Generate time to next event, \( s_1 \), and define \( t_1 = t_0 + s_1 \).

viii. Return to top of loop to generate the next event and inter-event time \( s_2 \) and define \( t_2 = t_1 + s_2 \).
ix. Repeat loop as many times as desired.

To simulate inter-event times \( s_j \), one makes use of the fact that the inter-event time \( S \) is an exponentially distributed random variable, i.e.

\[
P(S \geq s) = e^{-Rs}
\]

Then to randomly choose an inter-event time one

vii. Generates a new p-RN \( Y \);

viii. Defines \( s := \left\lfloor \log_e(Y) \right\rfloor \);

ix. Updates \( t \mapsto t + s \).