

---

# STOCHASTIC PROCESSES

## STAT 150

---

Instructor: Benson Au

KELVIN LEE

UC BERKELEY

# Contents

<b>1</b>	<b>Probability Review</b>	<b>3</b>
1.1	Basic Definitions . . . . .	3
1.2	Overview . . . . .	3
1.3	Useful Properties . . . . .	4
1.4	Conditional Probability . . . . .	4
1.5	Random Variables . . . . .	5
1.5.1	Discrete Random Variables . . . . .	5
1.5.2	Indicator Random Variable . . . . .	5
1.5.3	Continuous Random Variables . . . . .	5
1.5.4	Cumulative Distribution Function (CDF) . . . . .	6
1.5.5	Expectation . . . . .	6
1.5.6	Variance . . . . .	6
1.5.7	Moments . . . . .	7
1.5.8	Joint Distribution . . . . .	7
1.5.9	Independence . . . . .	7
1.5.10	Linearity of Expectation . . . . .	7
1.5.11	Convolution . . . . .	8
1.5.12	Gamma Distribution . . . . .	9
1.5.13	Moment Generating Function . . . . .	9
1.6	Conditional Probability (Cont'd) . . . . .	10
1.6.1	Conditional Expectation . . . . .	11
<b>2</b>	<b>Random Sums</b>	<b>13</b>
2.1	Mean and Variance of Random Sums . . . . .	14
<b>3</b>	<b>Markov Chains</b>	<b>15</b>
3.1	Discrete-time Markov Chains . . . . .	15
3.1.1	$n$ -step transition probabilities . . . . .	15
3.2	First Step Analysis . . . . .	16
3.2.1	The General Absorbing Markov Chain . . . . .	17
3.3	Random Walk . . . . .	18
3.4	Branching Process . . . . .	19
3.4.1	Generating Functions . . . . .	21

# 1 Probability Review

## 1.1 Basic Definitions

**Definition 1.1.1** (Probability Space). A *probability space*  $(\Omega, \mathcal{F}, \mathbb{P})$  is a triple consisting of a set  $\Omega$  called the *sample space*, a set  $\mathcal{F} \subseteq \Omega$  satisfying certain closure properties, and a function  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  that assigns probabilities to events in a coherent way.

**Requirements for  $\mathcal{F}$ :**

- (i)  $\Omega \in \mathcal{F}$ .
- (ii) If  $E \in \mathcal{F}$ , then  $E^c \in \mathcal{F}$ .
- (iii) If  $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ , then

$$\bigcap_{i=1}^{\infty} E_i \in \mathcal{F}.$$

**Requirements for  $\mathbb{P}$ :**

- (i)  $\mathbb{P}(\Omega) = 1$ .
- (ii) If  $\{E_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$  are pairwise disjoint (meaning  $E_i \cap E_j = \emptyset$  for  $i \neq j$ ), then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

**Definition 1.1.2** (Random Variable). A *random variable* is a function  $X : \Omega \rightarrow \mathbb{R}$  such that  $X^{-1}(B) \in \mathcal{F}$  whenever  $B$  is a "nice" subset of  $\mathbb{R}$ .

**Example 1.1.3.**  $\Omega = \{H, T\}$ ,  $\mathcal{F} = 2^{\Omega}$ ,  $\mathbb{P}(\{H\}) = \frac{1}{2}$ .  $X(H) = 1$ ,  $X(T) = 0$ .

$$\mathbb{P}(X = 1) = \mathbb{P}(\{H\}) = \frac{1}{2}, \quad \mathbb{P}(X = 0) = \mathbb{P}(\{T\}) = \frac{1}{2}.$$

## 1.2 Overview

**Definition 1.2.1** (Stochastic Process). A *stochastic process* is a collection  $\{X_t : t \in T\}$  of random variables  $X_t : \Omega \rightarrow S \subseteq \mathbb{R}$  all defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Here  $T$  is some index set (typically representing time) and  $S$  is the *state space*. One write this as

$X : \Omega \times T \rightarrow S$ ,  $(\omega, t) \mapsto X_t(\omega)$ . For a given outcome  $\omega \in \Omega$ , we get a sample path trajectory  $X(\omega) : T \rightarrow S, t \mapsto X_t(\omega)$ . A stochastic process can then be thought of as a random function.

The theme of this course is what can we say about the distribution of trajectories?

**Example 1.2.2** (Branching Process (DTDS)).  $X_0 = 1$ , one individual in the 0th generation individuals produce a random number of offspring, i.i.d.  $(\xi_i^{(n)})_{i \in \mathbb{N}, n \in \mathbb{N}_0}$ .

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i^{(n)}.$$

One interesting question would be what is  $\mathbb{P}(X_n = 0 \text{ eventually})$ , the probability of dying out?

**Example 1.2.3** (Poisson Process (CTDS)). Recall that the Poisson distribution is used to model the number of occurrences of a rare event in some fixed period of time. The Poisson process  $(N_t)_{t \geq 0}$  models the number of occurrences throughout time.  $N_t = \#$  of occurrences by time  $t$ .

## 1.3 Useful Properties

(i) (*DeMorgan*)

$$(E \cup F)^c = E^c \cap F^c, \quad (E \cap F)^c = E^c \cup F^c.$$

(ii) (*Complementation*)

$$\mathbb{P}(E) = 1 - \mathbb{P}(E^c).$$

(iii) (*Inclusion-exclusion*)

$$\begin{aligned} \mathbb{P}(E \cup F) &= \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) \\ \mathbb{P}\left(\bigcup_{i=1}^n E_i\right) &= \sum_{j=1}^n (-1)^{j-1} \sum_{S \in [n]: |S|=j} \mathbb{P}\left(\bigcap_{i \in S} E_i\right). \end{aligned}$$

(iv) (*Partitioning*) If  $\bigsqcup_{i=1}^{\infty} E_i = \Omega$ , then

$$\mathbb{P}(F) = \mathbb{P}\left(\bigsqcup_{i=1}^{\infty} (F \cap E_i)\right) = \sum_{i=1}^{\infty} \mathbb{P}(F \cap E_i)$$

## 1.4 Conditional Probability

**Conditioning:** For  $\mathbb{P}(F) > 0$ ,

$$\mathbb{P}(E | F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

$\mathbb{P}(\cdot | F)$  defines a new probability measure on  $(\Omega, \mathcal{F})$ .

**Multiplication rule:**

$$\mathbb{P}(E \cap F) = \mathbb{P}(F) \mathbb{P}(E | F).$$

If  $\bigcup_{i=1}^{\infty} F_i = \Omega$ , then

$$\mathbb{P}(E) = \sum_{i=1}^{\infty} \mathbb{P}(E \cap F_i) = \sum_{i=1}^{\infty} \mathbb{P}(F_i) \mathbb{P}(E | F_i).$$

Bayes' rule:

$$\mathbb{P}(F_j | E) = \frac{\mathbb{P}(F_j) \mathbb{P}(E | F_j)}{\sum_{i=1}^{\infty} \mathbb{P}(F_i) \mathbb{P}(E | F_i)}$$

## 1.5 Random Variables

### 1.5.1 Discrete Random Variables

If  $X : \Omega \rightarrow S \subseteq \mathbb{R}$  is discrete,

$$\mathbb{P}(X \in E) = \sum_{x \in E} \mathbb{P}(X = x) = \sum_{x \in E \cap S} \mathbb{P}(X = x).$$

### 1.5.2 Indicator Random Variable

$$X(\omega) = \mathbb{I}_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E. \end{cases}$$

#### 1.5.2.1 Binomial Random Variable

$$X = \sum_{i=1}^n \mathbb{I}_{E_i}, \quad \mathbb{P}(E_i) = p.$$

$$p_X(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

### 1.5.3 Continuous Random Variables

If  $X$  continuous,

$$\mathbb{P}(X \in E) = \int_E f_X(x) dx.$$

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x) dx.$$

#### 1.5.3.1 Exponential Random Variable

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{x \geq 0}.$$

### 1.5.3.2 Gaussian Random Variable

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

### 1.5.4 Cumulative Distribution Function (CDF)

$$F_X : \mathbb{R} \rightarrow [0, 1],$$

$$F_X(r) = \mathbb{P}(X \leq r) = \mathbb{P}(X \in (-\infty, r]).$$

If  $X$  is discrete,

$$F_X(r) = \sum_{x_i \leq r} p_X(x_i).$$

If  $X$  is continuous,

$$F_X(r) = \int_{-\infty}^r f_X(x) dx.$$

$$\frac{d}{dr} F_X(r) = f_X(r).$$

### 1.5.5 Expectation

#### 1.5.5.1 Discrete case

$$\mathbb{E}[X] = \sum_{x_i \in S} x_i \mathbb{P}(X = x_i).$$

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n).$$

$$\mathbb{E}[g(X)] = \sum_{x_i \in S} g(x_i) \mathbb{P}(X = x_i).$$

#### 1.5.5.2 Continuous case

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \mathbb{P}(X \geq x) dx.$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

### 1.5.6 Variance

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

### 1.5.7 Moments

$$\mathbb{E}[X^m] = \int_0^\infty mx^{m-1}\mathbb{P}(X \geq x)dx.$$

### 1.5.8 Joint Distribution

#### 1.5.8.1 Discrete

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$$

#### 1.5.8.2 Continuous

$$\mathbb{P}((X,Y) \in E) = \int \int_E f_{X,Y}(x,y)dx dy$$

#### 1.5.8.3 Marginal Distribution

$$p_X(x) = \sum_{y \in S_Y} p_{X,Y}(x,y).$$

$$f_X(x) = \int_{y \in S_y} f_{X,Y}(x,y)dy$$

### 1.5.9 Independence

$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

$$\mathbb{P}(X \leq x, Y \leq y) = F_X(x)F_Y(y).$$

### 1.5.10 Linearity of Expectation

$$\mathbb{E} \left[ \sum_{i=1}^n c_i X_i \right] = \sum_{i=1}^n c_i \mathbb{E}[X_i]$$

If  $(X_i)_{i=1}^n$  independent,

$$(g(X_i))_{i=1}^n$$

independent.

$$\mathbb{E} \left[ \prod_{i=1}^n g(X_i) \right] = \prod_{i=1}^n \mathbb{E} [g(x_i)]$$

$$\text{Var} \left( \sum_{i=1}^n x_i \right) = \sum_{i=1}^n \text{Var}(x_i)$$

In general,

$$\text{Var} \left( \sum_{i=1}^n x_i \right) = \sum_{i,j=1}^n \text{Cov}(x_i, x_j)$$

### 1.5.11 Convolution

**Discrete case:**  $X, Y$  discrete  $X \perp\!\!\!\perp Y$

$$\begin{aligned} \mathbb{P}(X + Y = z) &= \sum_Y \mathbb{P}(X + Y = z, Y = y) \\ &= \sum_y \mathbb{P}(X = z - y, Y = y) \\ &= \sum_y \mathbb{P}(X = z - y) \mathbb{P}(Y = y) \quad (= \sum_x \mathbb{P}(X = x) \mathbb{P}(Y = z - x)). \end{aligned}$$

If  $X, Y$  are  $\mathbb{Z}$ -valued, this becomes

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \sum_{k=-\infty}^{\infty} \mathbb{P}(X = n - k) \mathbb{P}(Y = k) \\ &= \sum_{k=-\infty}^{\infty} \mathbb{P}(X = k) \mathbb{P}(Y = n - k) \\ &= (\mathbb{P}_X * \mathbb{P}_Y)(n) \end{aligned}$$

**Example 1.5.1** (Poisson).  $X \sim \text{Poisson}(\lambda)$ ,  $Y \sim \text{Poisson}(\mu)$ ,  $X + Y \sim \text{Poisson}(\lambda + \mu)$

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k) \\ &= \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda^k \mu^{n-k} \\ &= \frac{e^{-(\lambda+\mu)}}{n!} (\lambda + \mu)^n \\ &= \mathbb{P}(Z = n) \end{aligned}$$

where  $Z \sim \text{Poisson}(\lambda + \mu)$ .

**Continuous case:**  $X, Y$  continuous

$$\begin{aligned} \mathbb{P}(X + Y \leq z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^z f_X(x) f_Y(y - x) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x) f_Y(y - x) dx dy \end{aligned}$$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx = \int_{-\infty}^{\infty} f_Y(y) f_X(z - y) dy = f_X * f_Y.$$



**Example 1.5.2** (Convolution in uniform distributions).  $X, Y \sim U[0, 1]$ ,  $X \perp\!\!\!\perp Y$ .

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx.$$

$$f_X(x) = \mathbb{I}_{[0,1]}(x) \quad f_Y(y) = \mathbb{I}_{[0,1]}(y)$$

so

$$\begin{aligned} f_{X+Y}(z) &= \int_{x \in [0,1], z-x \in [0,1]} 1dx \\ &= \int_{x \in [0,1], x \in [-1+z, z]} 1dx \\ &= \int_{\max(0, -1+z)}^{\min(1, z)} 1dx \\ &= \min(1, z) - \max(0, -1+z). \end{aligned}$$

### 1.5.12 Gamma Distribution

**Definition 1.5.3** (Gamma function). Let  $\alpha > 0$ . The *gamma function*  $\Gamma : (0, \infty) \rightarrow (0, \infty)$  is defined by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx = \mathbb{E}[X^{\alpha-1}]$$

where  $X \sim \text{Exp}(1)$  Let  $\alpha, \lambda > 0$ . The  $\text{Gamma}(\alpha, \lambda)$  distribution is defined by

$$f(x) = \frac{\lambda^{\alpha} x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} \mathbb{I}_{x \geq 0}.$$

**Exercise 1.5.4.**  $\Gamma(n) = (n-1)!$  for  $n \in \mathbb{N}$ . (Hint: use induction)

### 1.5.13 Moment Generating Function

**Definition 1.5.5** (MGF). For a random variable  $X$ , the *moment generating function* (MGF) is the function  $M_X : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$ ,

$$M_X(t) = \mathbb{E}[e^{tX}].$$

If  $M_X(t) < +\infty$  for  $t \in (-\epsilon, \epsilon)$ , then

$$M_x(t) = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}[x^k]}{k!} \text{ for } |t| < \epsilon$$

For independent RVs  $(X_i)_{i=1}^n$ ,

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t)$$

**Exercise 1.5.6.** If  $X \sim \text{Exp}(\lambda)$ , then  $M_x(t) = \frac{\lambda}{\lambda-t}$  if  $t < \lambda$ ,  $+\infty$  otherwise.

If  $X \sim \text{Gamma}(n, \lambda)$ , then

$$M_x(t) = \left( \frac{\lambda}{\lambda-t} \right)^n.$$

If  $X \sim \text{Gamma}(\alpha, \lambda)$ , then

$$M_x(t) = \left( \frac{\lambda}{\lambda-t} \right)^{\alpha}.$$

## 1.6 Conditional Probability (Cont'd)

**Exercise 1.6.1** (Generalization).  $(X_i)_{i=1}^n, (Y_j)_{j=1}^m$

$$p_{X_1, \dots, X_n | Y_1, \dots, Y_m} (x_1, \dots, x_n | y_1, \dots, y_m) = ?$$

**Example 1.6.2.** Let  $M \in \mathbb{N}$  and  $p, q \in (0, 1)$ . Consider  $N \sim \text{Bin}(M, q)$  and  $X \sim \text{Bin}(N, p)$ . What is the distribution of  $X$ ?

$$\begin{aligned} \mathbb{P}(X = k) &= \sum_{n=0}^M \mathbb{P}(N = n) \mathbb{P}(X = k | N = n) \\ &= \sum_{n=0}^M \binom{M}{n} q^n (1-q)^{M-n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{p^k}{k!} \sum_{n=k}^M \frac{M!}{(M-n)!(n-k)!} q^n (1-q)^{M-n} (1-p)^{n-k} \\ &= \frac{p^k}{k!(M-k)!} \sum_{n=k}^M \frac{M!(M-k)!}{(M-n)!(n-k)!} q^n (1-q)^{M-n} (1-p)^{n-k} \\ &= \binom{M}{k} p^k q^k \sum_{n=k}^M \binom{M-k}{n-k} q^{n-k} (1-q)^{M-j} (1-p)^{n-k} \\ &= \binom{M}{k} p^k q^k \sum_{t=0}^{M-k} \binom{M-k}{t} (q(1-p))^t (1-q)^{M-k-t} \\ &= \binom{M}{k} (pq)^k (q(1-p) + (1-q))^{M-k} \\ &= \binom{M}{k} (pq)^k (1-pq)^{M-k}. \end{aligned}$$

Thus,  $X \sim \text{Bin}(M, pq)$ .

**Remark.** What if  $k > n$  in  $\mathbb{P}(X = k | N = n)$  above in the first line? The probability is simply 0.

**Question.** Why does this answer make sense?

**Answer.** Think about retesting whenever we succeeded for the first  $M$  trials. Then  $X$  is simply the number of trials with double successes, thus we have the  $pq$  parameter.

**Exercise 1.6.3.** Consider  $N \sim \text{Poisson}(\lambda)$ ,  $X \sim \text{Bin}(N, p)$ . What is the distribution of  $X$ ?

**Answer.**  $X \sim \text{Poisson}(\lambda p)$ .

**Question.** How can we interpret this?

**Answer.** We can interpret  $X$  as the number of customers visiting a store who purchase something.

### 1.6.1 Conditional Expectation

For  $X, Y$  discrete,  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Assume  $\mathbb{E}[|g(X)|] = \sum_x |g(x)p_X(x)| < \infty$ .

**Definition 1.6.4** (Conditional expectation). The *conditional expectation* is defined as

$$\mathbb{E}[g(X) | Y = y] = \sum_x g(x)p_{X|Y}(x|y)$$

if  $p_Y(y) \neq 0$ .

**Remark.** Note that  $\mathbb{E}[g(X) | Y = y]$  is a real number, whereas  $\mathbb{E}[g(X) | Y]$  is a random variable.

#### 1.6.1.1 Tower Property

$$\begin{aligned} \mathbb{E}[\mathbb{E}[g(X) | Y]] &= \mathbb{E}\left[\sum_y \mathbb{E}[g(X) | Y = y]p_Y(y)\right] \\ &= \sum_y \mathbb{E}[g(X) | Y = y]p_Y(y) \\ &= \sum_y \sum_x g(x)p_{X|Y}(x|y)p_Y(y) \\ &= \sum_x g(x) \sum_y p_{X|Y}(x|y)p_Y(y) \\ &= \sum_x g(x)p_X(x) \\ &= \mathbb{E}[g(X)]. \end{aligned}$$

**Remark.** One intuitive example would be considering the averages of heights of students from a classroom. We divide it into several groups and let  $Y$  denote the whichever group we select and let  $\mathbb{E}[g(X) | Y]$  be the average of those from group  $Y$ . Then the average height of the entire classroom  $\mathbb{E}[g(X)]$  is equivalent to the average of the average of heights of each group, which is  $\mathbb{E}[\mathbb{E}[g(X) | Y]]$ .

#### Properties of conditional expectations:

1.  $\mathbb{E}[c_1g(x_1) + c_2h(x_2) | Y = y] = c_1\mathbb{E}[g(X_1) | Y = y] + c_2\mathbb{E}[h(X_2) | Y = y]$
2. If  $g \geq 0$ , then  $\mathbb{E}[g(x) | Y = y] \geq 0$ .
3.  $\mathbb{E}[f(X, Y) | Y = y] = \mathbb{E}[f(X, y) | Y = y]$ .
4. If  $X \perp\!\!\!\perp Y$ ,  $\mathbb{E}[g(X) | Y = y] = \mathbb{E}[g(X)]$
5.  $\mathbb{E}[g(x)h(y) | Y = y] = h(y)\mathbb{E}[g(x) | Y = y]$
6.  $\mathbb{E}[g(x)h(y)] = \sum_y h(y)\mathbb{E}[g(x) | Y = y]p_Y(y) = \mathbb{E}[h(Y)\mathbb{E}[g(X) | Y]]$

*Proof of 3.*

$$\begin{aligned}\mathbb{E}[f(X, Y) \mid Y = y] &= \sum_{x, z} f(x, z) p_{X, Y \mid Y}(x, z \mid y) \\ &= \sum_{x, z} f(x, z) \frac{p_{X, Y, Y}(x, z, y)}{p_Y(y)} \\ &= \sum_x f(x, y) \frac{p_{X, Y}(x, y)}{p_Y(y)} \\ &= \mathbb{E}[f(X, y) \mid Y = y].\end{aligned}$$

□

**Remark.**  $\mathbb{E}[f(X, y)] \neq \mathbb{E}[f(X, y) \mid Y = y]$ .

## 2 Random Sums

**Definition 2.0.1.** Let  $(\xi_i)_{i=1}^{\infty}$  be i.i.d random variables,  $N$  be a  $\mathbb{N}_0$ -valued random variable,  $N \perp\!\!\!\perp (\xi_i)_{i=1}^{\infty}$ . The *random sum* is defined as

$$X = \sum_{i=1}^N \xi_i = \sum_{n=0}^{\infty} \left( \sum_{i=1}^n \xi_i \right) \mathbf{1}_{N=n} = \begin{cases} \sum_{i=1}^n \xi_i & \text{if } N = n \geq 1 \\ 0 & \text{if } N = 0. \end{cases}$$

**Question.** What is the distribution of  $X$ ?

Let  $X, N$  be random variables.  $N$  is  $\mathbb{N}_0$ -valued. The condition CDF is

$$F_{X|N}(x|n) = \mathbb{P}(X \leq x \mid N = n)$$

if  $\mathbb{P}(N = n) \neq 0$ . This is an actual CDF, but for the random variable  $X \mid N = n$ .

Suppose that  $X$  is continuous and  $F_{X|N}(x|n)$  is a differentiable function of  $x$  for each  $n$  such that  $p_N(n) > 0$ . The conditional PDF is

$$f_{X|N}(x|n) = \frac{d}{dx} F_{X|N}(x|n).$$

$$\begin{aligned} \int_a^b f_{X|N}(x|n) dx &= F_{X|N}(b|n) - F_{X|N}(a|n) \\ &= \mathbb{P}(X \in [a, b] \mid N = n). \end{aligned}$$

**Answer.**

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{n=0}^{\infty} \mathbb{P}(N = n) \mathbb{P}(X \leq x \mid N = n).$$

$$f_X(x) = \sum_{n=0}^{\infty} \mathbb{P}(N = n) f_{X|N}(x|n).$$

## 2.1 Mean and Variance of Random Sums

Assume  $\mathbb{E}[N] = \nu$  and  $\mathbb{E}[\xi_i] = \mu$ . Then

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X \mid N]] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N \xi_i \mid N\right]\right] \\ &= \mathbb{E}[N\mathbb{E}[\xi_1]] \\ &= \mathbb{E}[N\mu] \\ &= \mu\nu.\end{aligned}$$

## 3 Markov Chains

### 3.1 Discrete-time Markov Chains

**Definition 3.1.1** (Markov process). A is a stochastic process  $(X_t)_{t \in T}$  such that the future, given the present, is independent of the past.

**Definition 3.1.2** (Markov property). The *Markov property* for a DTDS stochastic process is

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

**Example 3.1.3** (Gambler's ruin).  $(X_n)_{n=0}^\infty$ ,  $X_n$  = your wealth after  $n$  turns. Stop if  $X_n = 0$  or 5. Each play, you win \$1 with probability  $p$  and lose \$1 with probability  $1 - p$  independently of all previous plays. This process satisfies the markov property.

**Example 3.1.4** (Ehrenfest model). Box of  $N$  particles.  $X_n$  = number of particles on the left side at time  $n$ .  $N - X_n$  be the number of particles on the other side.

$$\begin{aligned}\mathbb{P}(X_{n+1} = i + 1 \mid X_n = i) &= \frac{N - i}{N} \\ \mathbb{P}(X_{n+1} = i - 1 \mid X_n = i) &= \frac{i}{N}.\end{aligned}$$

#### **Theorem 3.1.5.**

Joint PMF of the Markov Chain is determined by initial distribution and  $P = (p_{i,j})_{i,j \in S}$ .

*Proof.*

$$\begin{aligned}\mathbb{P}(X_n = i_n, \dots, X_0 = i_0) &= \mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \mathbb{P}(X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= p_{i_{n-1}, i_n} p_{i_{n-2}, i_{n-1}} \cdots p_{i_0, i_1} \mathbb{P}(X_0 = i_0).\end{aligned}$$

□

#### 3.1.1 $n$ -step transition probabilities

$$p_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

**Theorem 3.1.6.**

$$p_{i,j}^{(m)} = \mathbb{P}(X_{n+m} = j \mid X_n = i) = (P^m)_{i,j}.$$

*Proof.*

$$\begin{aligned} \mathbb{P}(X_{n+m+1} = j \mid X_n = i) &= \sum_{k \in S} \mathbb{P}(X_{n+m+1} = j, X_{n+m} = k \mid X_n = i) \\ &= \sum_{k \in S} \mathbb{P}(X_{n+m+1} = j \mid X_{n+m} = k) \mathbb{P}(X_{n+m} = k \mid X_n = i). \end{aligned}$$

□

**Example 3.1.7.**

$$\begin{aligned} p_{i,j}^{(2)} &= \mathbb{P}(X_2 = j \mid X_0 = i) \\ &= \sum_{k \in S} \mathbb{P}(X_2 = j, X_1 = k \mid X_0 = i) \\ &= \sum_{k \in S} \mathbb{P}(X_2 = j \mid X_1 = k, X_0 = i) \mathbb{P}(X_1 = k \mid X_0 = i) \\ &= \sum_{k \in S} P_{i,k} P_{k,j} \\ &= (P^2)_{i,j} \end{aligned}$$

**Example 3.1.8** (Inventory model).  $X_n$  = inventory that you have of this product after the  $n$ th business day. If  $X_n \leq s$ , place an order that brings inventory back to  $S$  by next morning.  $\xi_n$  = demand on  $n$ th day and  $(\xi_n)$  are i.i.d..

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \mathbb{P}(\xi_{n+1} = S - j) & \text{if } i \leq s \\ \mathbb{P}(\xi_{n+1} = i - j) & \text{if } i > s. \end{cases}$$

$\lim_{n \rightarrow \infty} \mathbb{P}(X_n < 0)$  = chance of excess demand.

## 3.2 First Step Analysis

Consider  $(X_n)_{n \geq 0}$  Markov chain on  $\{1, \dots, r\} \cup \{r+1, \dots, N\}$  where  $\{1, \dots, r\}$  are the *transient states* and  $\{r+1, \dots, N\}$  are the *absorbing states* such that

$$\begin{aligned} \lim_{n \rightarrow \infty} p_{i,j}^{(n)} &= 0 & \forall i, j \in \{1, \dots, r\} \\ \lim_{n \rightarrow \infty} p_{i,i}^{(n)} &= 1 & \forall i \in \{r+1, \dots, N\} \end{aligned}$$

Then we can express the transition matrix  $P$  as

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix},$$



where  $Q$  and  $R$  is some transition matrices for the corresponding partitioned states and  $0$  is the zero matrix and  $I$  is the identity matrix.

Let  $T = \min \{n \geq 0 : X_n \geq r + 1\}$  be the time of absorption and  $X_T$  be the state we get absorbed into. Define  $u_{i,k} = \mathbb{P}(X_T = k \mid X_0 = i)$ . Then we have

$$\begin{aligned} u_{i,k} &= \sum_{j=1}^N \mathbb{P}(X_T = k, X_1 = j \mid X_0 = i) \\ &= \sum_{j=1}^N \mathbb{P}(X_T = k \mid X_1 = j, X_0 = i) \mathbb{P}(X_1 = j \mid X_0 = i) \\ &= \sum_{j=1}^N p_{i,j} \mathbb{P}(X_T = k \mid X_1 = j) \\ &= \sum_{j=1}^N \mathbb{P}(X_T = k \mid X_0 = j) \\ &= \sum_{j=1}^r p_{i,j} u_{j,k} + \sum_{j=r+1, j \neq k}^N p_{i,j} u_{j,k} + p_{i,k} u_{k,k}. \end{aligned}$$

Thus,

$$u_{i,k} = \sum_{j=1}^r P_{i,j} u_{j,k} + p_{i,k}$$

Hence, we have

$$U = QU + R \implies (I - Q)U = R \implies U = (I - Q)^{-1}R,$$

where  $U$  contains all the  $(u_{i,k})_{i \in \{1, \dots, r\}, k \in \{r+1, \dots, N\}}$ .

### 3.2.1 The General Absorbing Markov Chain

Let's suppose that associated with each transient state  $i$  is a rate  $g(i)$  and that we wish to determine the mean total rate that is accumulated up to absorption. Let  $v_i$  be this mean total amount, where the subscript  $i$  denotes the starting position  $X_0 = i$ , i.e.,

$$v_i = \mathbb{E} \left[ \sum_{n=0}^{T-1} g(X_n) \mid X_0 = i \right]$$

The choice  $g = 1$  will give  $v_i = \mathbb{E}[T \mid X_0 = i]$ . We can also write for  $i \in \{1, \dots, r\}$  that

$$\begin{aligned} v_i &= g(i) + \mathbb{E} \left[ \sum_{n=1}^{T-1} g(X_n) \mid X_0 = i \right] \\ &= g(i) + \sum_{j=1}^N p_{i,j} v_j \quad (= \sum_{j=1}^N p_{i,j} (g(i) + v_i)). \end{aligned}$$

Then we can condense this into the following form

$$v = g + Qv \implies v = (I - Q)^{-1}g.$$

where  $v = (v_i)_{i \in \{1, \dots, r\}}$  and  $g = (g(i))_{i \in \{1, \dots, r\}}$ .

### 3.3 Random Walk

$(\xi_n)_{n=1}^\infty$  i.i.d and  $\mathbb{Z}$ -valued. Then

$$X_n = \sum_{i=0}^n \xi_i.$$

$$\begin{aligned} \mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1) &= \mathbb{P}(\xi_{n+1} = j - i \mid \xi_n = i - i_{n-1}, \dots, \xi_1 = i_1) \\ &= \mathbb{P}(\xi_{n+1} = j - i) \\ &= \mathbb{P}(\xi_{n+1} = j - i \mid X_n = i). \end{aligned}$$

**Example 3.3.1** (Gambler's Ruin). Win 1 dollar with probability  $p$  and lose 1 dollar with probability  $q = 1 - p$ . Stop when we lose all money or make  $N$  dollars. We are interested in  $u_k = \mathbb{P}(X_T = 0 \mid X_0 = k)$  and  $v_k = \mathbb{E}[T \mid X_0 = k]$ . Clearly  $u_0 = 1, u_N = 0$ . For  $k = 1, \dots, N-1$ , we have

$$u_k = pu_{k+1} + qu_{k-1} \implies q(u_k - u_{k-1}) = p(u_{k+1} - u_k)$$

Let  $\Delta_{k+1} = u_{k+1} - u_k$ . Then we have

$$\begin{aligned} q\Delta_k &= p\Delta_{k+1} \\ \Delta_{k+1} &= \frac{q}{p}\Delta_k = \dots = \left(\frac{q}{p}\right)^k \Delta_1. \\ \sum_{i=1}^m \Delta_i &= \Delta_1 \sum_{i=1}^m \left(\frac{q}{p}\right)^{i-1} = \sum_{i=1}^m u_i - u_{i-1} = u_m - u_0 = u_m - 1 \end{aligned}$$

Thus,

$$u_m = 1 + \Delta_1 \frac{1 - \left(\frac{q}{p}\right)^m}{1 - \frac{q}{p}} \quad m = 1, \dots, N$$

When  $m = N$ ,

$$0 = 1 + \Delta_1 \frac{1 - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} \implies \Delta_1 = -\frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}.$$

Substituting the expression for  $\Delta_1$  gives

$$u_m = 1 + \left( -\frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N} \right) \left( \frac{1 - \left(\frac{q}{p}\right)^m}{1 - \frac{q}{p}} \right) = 1 - \frac{1 - \left(\frac{q}{p}\right)^m}{1 - \left(\frac{q}{p}\right)^N} = \frac{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}.$$

Note that  $p \neq q$ . If  $p = q$ , then

$$\sum_{i=1}^m \Delta_i = \Delta_1 m = u_m - 1 \implies u_m = \frac{N - m}{N}.$$

If we take limit as  $N \rightarrow \infty$  for  $p \leq q$ , then

$$\lim_{N \rightarrow \infty} u_m = 1,$$

which implies that we will be broke at the end no matter how much money we started with. If  $p > q$ , then

$$\lim_{N \rightarrow \infty} u_m = \left(\frac{q}{p}\right)^m.$$

If  $m$  is large, then this quantity becomes small. This implies that if  $p > q$  and we started with a lot of money, then the chance of us being broke ultimately becomes smaller.

Now let's compute  $v_k$  when  $p = q = \frac{1}{2}$ . Clearly,  $v_0 = 0$  and  $v_N = 0$ . For  $k = 1, \dots, N-1$ , we have

$$v_k = 1 + \frac{1}{2}v_{k+1} + \frac{1}{2}v_{k-1}.$$

Let  $\Delta_k = v_k - v_{k-1}$ . Then we have

$$0 = 1 + \frac{1}{2}(\Delta_{k+1} - \Delta_k).$$

Summing both sides gives

$$\sum_{k=1}^m 0 = m + \sum_{k=1}^m \frac{1}{2}(\Delta_{k+1} - \Delta_k) \implies \Delta_1 = 2m + \Delta_{m+1} \quad m = 0, \dots, N-1.$$

Then

$$\begin{aligned} \sum_{m=0}^k \Delta_1 &= \sum_{m=0}^k (2m + \Delta_{m+1}) \\ (k+1)\Delta_1 &= (k+1)v_1 = \sum_{m=0}^k 2m + \sum_{m=0}^k \Delta_{m+1} \\ (k+1)v_1 &= k(k+1) + (v_{k+1} - v_0) \implies (k+1)v_1 = k(k+1) + v_{k+1}. \end{aligned}$$

Take  $k = N-1$  gives

$$Nv_1 = (N-1)N + 0 \implies v_1 = N-1.$$

Then

$$v_{k+1} = (k+1)(v_1 - k) = (k+1)(N-1-k).$$

Hence,

$$v_k = k(N-k).$$

### 3.4 Branching Process

$(\xi_i^{(n)})_{i=1, n=0}^{\infty, \infty}$  i.i.d.  $\mathbb{N}_0$ -valued random variables where  $\xi_i^{(n)}$  is the number of offspring of  $i$ th individual in  $n$ th generation.  $X_0 = 1$ .  $\mathbb{E}[\xi_i] = \mu$  and  $\text{Var}(\xi_i) = \sigma^2$ . The population of at time  $n+1$  is

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_i^{(n)}$$

Our goal is to compute  $\mathbb{P}(X_n = 0 \text{ eventually} \mid X_0 = 1)$ . But let's first compute  $\mathbb{E}[X_{n+1}]$  and  $\text{Var}(X_{n+1})$ . Recall that

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^N \xi_i\right] &= \mathbb{E}[N]\mathbb{E}[\xi_i] \\ \text{Var}\left(\sum_{i=1}^N \xi_i\right) &= \text{Var}(N)\mathbb{E}[\xi_i]^2 + \text{Var}(\xi_i)\mathbb{E}[N]. \end{aligned}$$

Then we have

$$\begin{aligned}\mathbb{E}[X_{n+1}] &= \mu \mathbb{E}[X_n] = \mu^{n+1} \\ \text{Var}(X_{n+1}) &= \mu^2 \text{Var}(X_n) + \mu^n \sigma^2.\end{aligned}$$

$$\begin{aligned}c_0 &= \text{Var}(X_0) = 0 \\ c_n &= \text{Var}(X_n) \\ c_{n+1} &= \mu^2 c_n + \mu^n \sigma^2.\end{aligned}$$

Define the generating function  $f(x)$  as

$$\begin{aligned}f(x) &= \sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty} c_{n+1} x^{n+1} = \mu^2 x \sum_{n=0}^{\infty} c_n x^n + \sigma^2 x \sum_{n=0}^{\infty} \mu^n x^n \\ &= \mu^2 x f(x) + \frac{\sigma^2 x}{1 - \mu x}.\end{aligned}$$

Then

$$f(x) = \frac{\sigma^2 x}{(1 - \mu x)(1 - \mu^2 x)} = \sigma^2 x \frac{1}{1 - \mu x} \frac{1}{1 - \mu^2 x}.$$

Since

$$\sum_{j=1}^{\infty} c_j x^j = \sigma^2 x \sum_{n=0}^{\infty} \mu^n x^n \cdot \sum_{m=0}^{\infty} \mu^{2m} x^m,$$

the coefficient of  $x^{j-1} = \sum_{k=0}^{j-1} x^k x^{j-1-k}$  is

$$c_j = \sum_{k=0}^{j-1} \mu^k \mu^{2(j-1-k)}.$$

Thus

$$\text{Var}(X_n) = \sigma^2 \mu^{n-1} \cdot \begin{cases} n & \text{if } \mu = 1 \\ \frac{1 - \mu^{n-1}}{1 - \mu} & \text{if } \mu \neq 1. \end{cases}$$

**Remark.** When  $\mu = 1$ , expectation is constant, variance is growing linearly. When  $\mu \neq 1$ , expectation is increasing/decreasing geometrically, same with variance.

Now let  $T = \min \{n \geq 0 : X_n = 0\}$  be the time the population dies out and let  $u_n = \mathbb{P}(T \leq n) = \mathbb{P}(X_n = 0)$ . Then  $\lim_{n \rightarrow \infty} u_n$  is the probability of extinction.

$$u_{n+1} = \sum_{k=0}^{\infty} p_k u_n^k$$

where  $p_k = \mathbb{P}(\xi = k)$ . We have  $u_0 = 0, u_1 = p_0$ .

Let  $\phi_\xi : [0, 1] \rightarrow [0, 1]$  be the generating function of  $\xi$  defined by

$$\phi_\xi(s) = \mathbb{E}[s^\xi] = \sum_{k=0}^{\infty} p_k s^k.$$

Then we have

$$u_{n+1} = \phi(u_n) \implies u_\infty = \lim_{n \rightarrow \infty} u_{n+1} = \lim_{n \rightarrow \infty} \phi(u_n) \implies u_\infty = \phi\left(\lim_{n \rightarrow \infty} u_n\right) = \phi(u_\infty).$$

Thus,  $u_\infty$  is a fixed point for  $\phi$ .

### 3.4.1 Generating Functions

Given any  $\mathbb{N}_0$ -valued random variable  $\xi$  with  $p_k = \mathbb{P}(\xi = k)$ . Then the generating function is given by

$$\phi_\xi(s) = \mathbb{E}[s^\xi] = \sum_{k=0}^{\infty} p_k s^k.$$

$\phi_\xi$  completely recovers the distribution of  $\xi$ . We have  $\phi_\xi(0) = p_0 \phi_\xi(1) = 1$ . We can recover  $p_k$  via

$$p_k = \frac{\phi^{(k)}(0)}{k!}.$$

Then

$$\mathbb{E}[X] = \phi'(1) = \sum_{k=1}^{\infty} k p_k.$$

In fact, one can check that

$$\begin{aligned}\phi''(1) &= \mathbb{E}[X(X-1)] \\ \phi^{(k)}(1) &= \mathbb{E}[X(X-1)\cdots(X-k+1)].\end{aligned}$$