

---

# 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 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  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  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 writes 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  $X$ ? Why does this answer make sense?

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

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.