

STAT 333

Stochastic Processes I

This page intentionally left blank.

Contents

1	Review of Probability	
1.1	Probability Spaces	2
1.2	Random Variables	3
1.3	Expectation	6
1.4	Joint Distributions	8
1.5	Independence	10
2	Conditional Distributions	
2.1	Jointly Discrete Case	14
2.2	Jointly Continuous Case	16
2.3	Conditioning	17

This page intentionally left blank.

1.

Review of Probability

-
- 1.1 Probability Spaces
 - 1.2 Random Variables
 - 1.3 Expectation
 - 1.4 Joint Distributions
 - 1.5 Independence
-

1.1 Probability Spaces

Probability Space

Def'n 1.1

A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ such that the following holds.

- (a) The **sample space** Ω is nonempty.
- (b) The **event space** \mathcal{F} is a σ -algebra on Ω . That is, $\mathcal{F} \subseteq 2^\Omega$ with the following properties:
 - (i) $\Omega \in \mathcal{F}$;
 - (ii) for every $A \in \mathcal{F}$, $(\Omega \setminus A) \in \mathcal{F}$; and *closure under complements*
 - (iii) for every countable $\{A_i\}_{i=1}^\infty \subseteq \mathcal{F}$, $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$. *closure under countable unions*
- (c) The **probability function** $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfies the following.
 - (i) For every countable and pairwise disjoint $\{A_i\}_{i=1}^\infty \subseteq \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mathbb{P}(A_i). \quad \sigma\text{-additivity}$$

- (ii) $\mathbb{P}(\Omega) = 1$.

(1.1) For simplicity, fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ throughout this section.

(1.2) A direct consequence of Def'n 1.1 is the following: for every $A \in \mathcal{F}$,
 Probability of the
 Complement of an Event

$$\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A).$$

Conditional Probability

Def'n 1.2

Let $A, B \in \mathcal{F}$ be such that $\mathbb{P}(B) \neq 0$. The **conditional probability** of A given B occurs, denoted as $\mathbb{P}(A|B)$, is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

(1.3) Let $A, B \in \mathcal{F}$ be such that $\mathbb{P}(B) \neq 0$.

- (a) Note that

$$\mathbb{P}(A|\Omega) = \frac{\mathbb{P}(A \cap \Omega)}{\mathbb{P}(\Omega)} = \frac{\mathbb{P}(A)}{1} = \mathbb{P}(A),$$

as expected.

- (b) By rearranging,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B). \quad \text{multiplication rule}$$

For any finite $\{A_i\}_{i=1}^n \subseteq \mathcal{F}$, we can generalize the multiplication rule as follows:

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}\left(A_i \mid \bigcap_{j=1}^i A_j\right), \quad \text{generalized multiplication rule}$$

provided that $\mathbb{P}\left(\bigcap_{j=1}^i A_j\right) \neq 0$ for all $i \in \{1, \dots, n\}$.

(EX 1.4)
Rolling a Fair Die

Suppose that we roll a fair six-sided die once. Let A denote the event of rolling a number less than 4 and let B denote the event of rolling an odd number. Given that the roll is odd, what is the probability that the number rolled is less than 4?

Answer. Note that we are trying to calculate $\mathbb{P}(A|B)$. By definition, $A = \{1, 2, 3\}, B = \{1, 3, 5\}$. So it follows that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}\{1, 3\}}{\mathbb{P}\{1, 3, 5\}} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}. \quad \triangleleft$$

Note that we are *implicitly* defining the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as $(\{1, \dots, 6\}, 2^{\{1, \dots, 6\}}, |\cdot|)$ for (EX 1.4).

Def'n 1.3

Independent Events

We say $A, B \in \mathcal{F}$ are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

Theorem 1.1
Law of Total Probability

Let $\mathcal{C} \subseteq \mathcal{F}$ be a countable partition of Ω . Then

$$\mathbb{P}(A) = \sum_{B \in \mathcal{C}} \mathbb{P}(A|B) \mathbb{P}(B)$$

for every $A \in \mathcal{F}$.

Corollary 1.1.1
Bayes' Formula

Consider the setting of Theorem 1.1. Then for every $C \in \mathcal{C}$,

$$\mathbb{P}(C|A) = \frac{\mathbb{P}(A|C) \mathbb{P}(C)}{\sum_{B \in \mathcal{C}} \mathbb{P}(A|B) \mathbb{P}(B)}.$$

1.2 Random Variables

Def'n 1.4

Random Variable

A **random variable** (or **rv** for short) X is a function of the form $X : \Omega \rightarrow \mathbb{R}$, where Ω is the sample space of a probability space.

Def'n 1.5

Discrete Random Variable

Let X be a random variable. When the image of X is countable, we say X is **discrete**. There are two important functions that are associated with X .

(a) We define the **probability mass function** (or **pmf** for short) for X , denoted as p_X , by

$$p_X(x) = \mathbb{P}\{X = x\} \quad \forall x \in \mathbb{R}.$$

(b) We define the **cumulative distribution function** (or **cdf** for short) for X , denoted as F_X , by

$$F_X(x) = \mathbb{P}\{X \leq x\} = \sum_{y \leq x} p_X(y) \quad \forall x \in \mathbb{R}.$$

(1.5) Let X be a discrete random variable.

- (a) Sometimes it is handy to have the **tail probability function** (or **tpf** for short) for X , denoted as \bar{F}_X : it is defined as

$$\bar{F}_X(x) = 1 - F(x) \quad \forall x \in \mathbb{R}.$$

- (b) Let S be the image of X . We can order the elements of S in the increasing order, so that $S = \{x_i\}_{i=1}^n$ if S is finite or $S = \{x_i\}_{i=1}^\infty$ if S is infinite, where $x_i < x_{i+1}$ for all i . Then note that we can *recover* the pmf p_X of X from F_X by

$$p_X(x_1) = F_X(x_1)$$

and

$$p_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

for every $i \geq 2$.

(1.6) **Bernoulli** A **Bernoulli trial** is a random trial with probability $p \in [0, 1]$ of being a *success* and probability $1 - p$ of being a *failure*. If we let $X = 1$ if the trial is successful and $X = 0$ if it fails, then X is said to be a **Bernoulli** random variable with parameter p , denoted as $X \sim B(p)$. Note that X has a pmf

$$p_X(x) = p^x (1 - p)^{1-x}$$

for all $x \in \{0, 1\}$.

(1.7) **Binomial** A **binomial random variable generalizes Bernoulli random variable**. Consider the case where we run $n \in \mathbb{N}$ *independent* Bernoulli trials, each with probability $p \in (0, 1]$, where we let X denote the number of successes. Then we say X is a **binomial** random variable with parameters n, p , denoted as $X \sim \text{BIN}(n, p)$. The pmf of X is given by

$$p_X(x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad [1.1]$$

for all $x \in \{0, \dots, n\}$. Note that $\binom{n}{x}$ is the *number of distinct x -subsets of a n -set*. Here are some remarks.

- (a) A $\text{BIN}(1, p)$ simplifies to become $B(p)$.
- (b) Note that [1.1] is even defined for $n = 0$, in which case $p_X(0) = 1$. Such a distribution is said to be *degenerate* at 0.

(1.8) **Negative Binomial** Suppose that we have independent Bernoulli trials, each with success probability $p \in (0, 1]$ required to observe $n \in \mathbb{N}$ successes. If we let X denote the number of trials needed, then X is a **negative binomial** random variable with parameters n, p , denoted as $X \sim \text{NB}_t(n, p)$. X has a pmf

$$p_X(x) = \binom{x-1}{n-1} p^n (1 - p)^{x-n} \quad [1.2]$$

for every $x \in \mathbb{N}, x \geq n$.

- (a) Note that the appearance of $\binom{x-1}{n-1}$ instead of $\binom{x}{n}$ in [1.2]; this is because the final trial (i.e. the n th trial) must always be a success.
- (b) Sometimes, a negative binomial distribution is alternatively defined as the number of *failures* observed to achieve n successes. If Y denotes such a random variable and $X \sim \text{NB}_t(n, p)$, then clearly $X = Y + n$, which implies

$$p_Y(y) = \binom{y+n-1}{n-1} p^n (1 - p)^y$$

for all $y \in \mathbb{N} \cup \{0\}$. We denote $Y \sim \text{NB}_f(n, p)$.

(1.9) Geometric A **geometric** random variable is a *special case of negative binomial*: that is, if $X \sim \text{NB}_t(1, p)$ for some $p \in (0, 1]$, then we say X is a geometric random variable with success probability p , denoted as $X \sim \text{GEO}_t(p)$.¹

(1.10) Discrete Uniform If a random variable X is *equally likely* to take on values in a finite set $\{a, a+1, \dots, b\}$ for some $a, b \in \mathbb{Z}, a \leq b$, then we say X is a **discrete uniform** random variable, denoted as $X \sim \text{DU}(a, b)$. X has a pmf

$$p_X(x) = \frac{1}{b-a+1}$$

for every $x \in \{a, a+1, \dots, b\}$.

(1.11) Hypergeometric If X denotes the number of success objects in $n \in \mathbb{N}$ draws *without replacement* from a finite set of size $N \in \mathbb{N}$ containing exactly $r \in \mathbb{N}$ success objects, then X is a **hypergeometric** random variable with parameters N, r, n , denoted as $X \sim \text{HG}(N, r, n)$. X has a pmf

$$p_X(x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

for all $x \in \{\max\{0, n-N+r\}, \dots, \min\{n, r\}\}$.

(1.12) Poisson A random variable is called **Poisson** with parameter $\lambda > 0$, denoted as $X \sim \text{POI}(\lambda)$, if

$$p_X(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad [1.3]$$

for all $x \in \mathbb{N} \cup \{0\}$. Note that [1.3] is even defined for $\lambda = 0$, in which case $p_X(0) = 1$ (i.e. X is degenerate at 0).

(EX 1.13) Approximating Binomial with Poisson Show that when $n \in \mathbb{N}$ is large and $p \in (0, 1]$ is small, $p_X \sim p_Y$ where $X \sim \text{BIN}(n, p), Y \sim \text{POI}(np)$.

Proof. Let $x \in \{0, \dots, n\}$. Then

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} \frac{(n)_x}{x!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x}.$$

where $\lambda = np$. Now note that $(n)_x \sim n^x$, $1 - \frac{\lambda}{n} \sim 1$, and $\left(1 - \frac{\lambda}{n}\right)^n \sim e^{-\lambda}$ since n is large and $p = \frac{\lambda}{n}$ is small. Hence

$$p_X(x) = \frac{\lambda^x}{x!} \frac{(n)_x}{x!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x} \sim e^{-\lambda} \frac{\lambda^x}{x!} = p_Y(x),$$

as required. ◁

Continuous Random Variable

Def'n 1.6

Let X be a random variable. We say X is **continuous** if there exists nonnegative $f_X : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\mathbb{P}\{X \in B\} = \int_{x \in B} f_X(x) \, dx$$

for all measurable $B \subseteq \mathbb{R}$, where f_X is called the **probability density function (pmf)** of X . We also define

¹Similar to negative binomial, we write $X \sim \text{GEO}_f(p)$ if $X \sim \text{NB}_f(1, p)$.

the **cumulative distribution function** $F_X : \mathbb{R} \rightarrow [0, 1]$ of X by

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \forall x \in \mathbb{R}.$$

(1.14) Let X be a continuous random variable. Then note that

$$f_X = F'_X$$

by the fundamental theorem of calculus.

(1.15) A random variable X is called a **uniform** random variable on an interval $(a, b) \subseteq \mathbb{R}$, denoted as $X \sim U(a, b)$ if

$$f_X(x) = \frac{1}{b-a} \quad \forall x \in (a, b).$$

(1.16) A random variable X is called **Beta** with parameters $m, n \in \mathbb{N}$, denoted as $X \sim \text{BETA}(m, n)$, if

$$f_X(x) = \frac{(m+n-1)!}{(m-1)!(n-1)!} x^{m-1} (1-x)^{n-1} \quad \forall x \in (0, 1).$$

(1.17) A random variable X is called **Erlang** with parameters $n \in \mathbb{N}, \lambda > 0$, denoted as $X \sim \text{ERLANG}(n, \lambda)$ if

$$f_X(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \quad \forall x > 0.$$

(1.18) A random variable X is called **exponential** with parameter $\lambda > 0$, denoted as $X \sim \text{Exp}(\lambda)$, if

$$f_X(x) = \lambda e^{-\lambda x} \quad \forall x > 0.$$

Note that $\text{ERLANG}(1, \lambda)$ simplifies to $\text{Exp}(\lambda)$.

1.3 Expectation

Expectation of a Random Variable

Def'n 1.7 Let X be a random variable. Then we define the **expectation** of X , denoted as $\mathbb{E}(X)$, by

$$\mathbb{E}(X) = \begin{cases} \sum_{x \in \mathbb{R}: p_X(x) > 0} x p_X(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

if exists.

n th Moment of a Random Variable

Def'n 1.8 Let X be a random variable. For any $n \in \mathbb{N} \cup \{0\}$, if $\mathbb{E}(X^n)$ exists, then it is called the **n th moment** of X .

Variance, Standard Deviation of a Random Variable

Def'n 1.9 Let X be a random variable. We define the **variance** of X , denoted as $\text{var}(X)$, by

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

We define the **standard deviation** (*stdev*) of X , denoted as $\text{sd}(X)$, by

$$\text{sd}(X) = \sqrt{\text{var}(X)}.$$

Theorem 1.2

Law of the Unconscious Statistician (LOTUS)

Let X be a random variable and let $g : \mathbb{R} \rightarrow \mathbb{R}$ be measurable. Then

$$\mathbb{E}(g(X)) = \begin{cases} \sum_{x \in \mathbb{R}: p_X(x) > 0} g(x) p_X(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}.$$

Corollary 1.2.1

Let X be a random variable and let $a, b \in \mathbb{R}$.

$$(a) \quad \mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$

$$(b) \quad \text{var}(aX + b) = a^2 \text{var}(X).$$

Moment Generating Function of a Random Variable

Def'n 1.10

Let X be a random variable. We define the **moment generating function** (*mgf*) of X , denoted as φ_X , by

$$\varphi_X(t) = \mathbb{E}(e^{tX}) \quad \forall t \in \mathbb{R}.$$

(1.19)

Note that

$$\varphi_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} \mathbb{E}(X^n) \frac{t^n}{n!},$$

implying that $\mathbb{E}(X^n)$ is the coefficient of $\frac{t^n}{n!}$ in the above series expansion. In particular,

$$\mathbb{E}(X^n) = \varphi_X^{(n)}(0)$$

for all $n \in \mathbb{N}$.

(1.20)

It is worth noting that a mgf *uniquely* determines the probability distribution of a random variable.

(EX 1.21)

Let $X \sim \text{BIN}(n, p)$, where $n \in \mathbb{N}$, $p \in (0, 1]$. Find φ_X and use it to calculate $\mathbb{E}(X)$.

Answer. Observe that, for every $t \in \mathbb{R}$,

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}(e^{tX}) = \sum_{x=0}^n e^{tx} p_X(x) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (e^t p)^x (1-p)^{n-x} = (e^t p + 1 - p)^n, \end{aligned}$$

where the last equality holds by the binomial theorem. It follows that

$$\mathbb{E}(X) = \varphi_X'(0) = \frac{d}{dt} (e^t p + 1 - p)^n \Big|_{t=0} = n(e^t p + 1 - p)^{n-1} e^t p \Big|_{t=0} = np.$$

<

1.4 Joint Distributions

Random Vector

Def'n 1.11

Let X_1, \dots, X_n be random variables. Then we call the n -tuple $\mathbf{X} = (X_1, \dots, X_n)$ a **random vector**.

(a) The **joint cdf** of \mathbf{X} , denoted as $F_{\mathbf{X}}$, is defined as

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}\{\mathbf{X} \leq \mathbf{x}\} \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

(b) When X_1, \dots, X_n is discrete, we say \mathbf{X} is **jointly discrete**, and define the **joint pmf** of \mathbf{X} , denoted as $p_{\mathbf{X}}$, by

$$p_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}\{\mathbf{X} = \mathbf{x}\} \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

(c) When there exists nonnegative $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$\mathbb{P}\{\mathbf{X} \in S\} = \int_S f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}$$

for every $S \subseteq \mathbb{R}^n$, then we say \mathbf{X} is **jointly continuous** and call $f_{\mathbf{X}}$ a **joint pdf** of \mathbf{X} .

(1.22)

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector.

(a) Note that, for every $i \in \{1, \dots, n\}$,

$$F_{X_i}(x_i) = F_{\mathbf{X}}\left(\underbrace{\infty, \dots, \infty, x_i, \infty, \dots, \infty}_{i\text{th position}}\right) \quad \forall x_i \in \mathbb{R}.$$

We call F_{X_i} the ***i*th marginal cdf** of \mathbf{X} .

(b) In case \mathbf{X} is jointly discrete, for every $i \in \{1, \dots, n\}$,

$$p_{X_i}(x_i) = \sum_{\substack{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \in \mathbb{R} \\ : p_{\mathbf{X}}(x_1, \dots, x_n) > 0}} p_{\mathbf{X}}(x_1, \dots, x_n) \quad \forall x_i \in \mathbb{R}.$$

We call p_{X_i} the ***i*th marginal pmf** of \mathbf{X} .

(c) In case \mathbf{X} is jointly continuous, each X_i is continuous, and for every $i \in \{1, \dots, n\}$,

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \underbrace{\int_{-\infty}^{x_i} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{i\text{th from right}} f_{\mathbf{X}}(t_1, \dots, t_n) \, dt_1 \cdots dt_n \quad \forall x_i \in \mathbb{R}.$$

We call f_{X_i} the ***i*th marginal pdf** of \mathbf{X} . It is worth noting that

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \dots, x_n).$$

Proposition 1.3

Let $\mathbf{X} = (X_1, \dots, X_n)$ be jointly continuous. Then for any injective C^1 $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with nowhere vanishing Jacobian determinant,

$$f_{g(\mathbf{X})}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_g(g^{-1}(\mathbf{y}))|^{-1} \quad \forall \mathbf{y} \in g^{-1}(\mathbb{R}^n).$$

Expectation of a Random Vector

Def'n 1.12

Let \mathbf{X} be a random vector. Then we define the **expectation** of \mathbf{X} , denoted as $\mathbb{E}(\mathbf{X})$, by

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n)).$$

Covariance of Two Random Variables

Def'n 1.13

Let X, Y be random variables. Then we define the **covariance** of X, Y , denoted as $\text{cov}(X, Y)$, by

$$\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

(1.23)
CovarianceLet X, Y be random variables. Note that

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

In particular, $\text{cov}(X, X) = \text{var}(X)$.Theorem 1.4
Multivariate LOTUSLet \mathbf{X} be a random vector and let $g : \mathbb{R}^n \rightarrow \mathbb{R}$. Then

$$\mathbb{E}(g(\mathbf{X})) = \begin{cases} \sum_{\mathbf{x} \in \mathbb{R}^n: p_{\mathbf{X}}(\mathbf{x}) > 0} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) & \text{if } \mathbf{X} \text{ is jointly discrete} \\ \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} & \text{if } \mathbf{X} \text{ is jointly continuous} \end{cases}.$$

Corollary 1.4.1
Linearity of ExpectationLet X_1, \dots, X_n be random variables and let $a_1, \dots, a_n \in \mathbb{R}$. Then

$$\mathbb{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i).$$

Corollary 1.4.2

Let X_1, X_2 be random variables and let $a_1, a_2 \in \mathbb{R}$. Then

$$\text{var}(a_1 X_1 + a_2 X_2) = a_1^2 \text{var}(X_1) + a_2^2 \text{var}(X_2) + 2a_1 a_2 \text{cov}(X_1, X_2).$$

Joint MGF of a Random Vector

Def'n 1.14

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector. We define the **joint mgf** of \mathbf{X} , denoted as $\varphi_{\mathbf{X}}$, by

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T \mathbf{X}} \quad \forall \mathbf{t} \in \mathbb{R}^n.$$

(1.24)
Joint MomentLet \mathbf{X} be a random vector. Then for every $i_1, \dots, i_n \in \mathbb{N} \cup \{0\}$,

$$\mathbb{E}\left(X_1^{i_1} \cdots X_n^{i_n}\right) = \left. \frac{\partial^{\sum_{j=1}^n i_j}}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}} \varphi_{\mathbf{X}}(\mathbf{x}) \right|_{\mathbf{x}=\mathbf{0}}.$$

1.5 Independence

Independent Random Variables

Def'n 1.15

Let X_1, \dots, X_n be random variables. We say X_1, \dots, X_n are **independent** if

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

for all $(x_1, \dots, x_n) \in \mathbb{R}^n$.

(1.25)

Let X_1, \dots, X_n be random variables and let $\mathbf{X} = (X_1, \dots, X_n)$.

- (a) If \mathbf{X} is jointly discrete, then Def'n 1.15 is equivalent to saying that $p_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i)$ for every $(x_1, \dots, x_n) \in \mathbb{R}^n$.
- (b) If \mathbf{X} is jointly continuous, then Def'n 1.15 is equivalent to saying that $f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$ for every $(x_1, \dots, x_n) \in \mathbb{R}^n$.
- (c) If $n = 2$ and X_1, X_2 are independent, note that $\text{cov}(X_1, X_2) = 0$. However the converse does not hold in general.

Theorem 1.5

MGF of the Sum of Independent
Random Variables

Let X_1, \dots, X_n be independent random variables. Then

$$\phi_{\sum_{i=1}^n X_i} = \prod_{i=1}^n \phi_{X_i}.$$

Corollary 1.5.1

Let X_1, \dots, X_n be iid random variables. Then

$$\phi_{\sum_{i=1}^n X_i} = \phi_{X_1}^n.$$

(EX 1.26)

Sum of Independent
Binomial Random Variables

Let $X_1 \sim \text{BIN}(n_1, p), \dots, X_m \sim \text{BIN}(n_m, p)$, where $n_1, \dots, n_m \in \mathbb{N}, p \in (0, 1]$. Find the distribution of $\sum_{i=1}^m X_i$.

Answer. Observe that, for every $t \in \mathbb{R}$,

$$\phi_{\sum_{i=1}^m X_i}(t) = \prod_{i=1}^m \phi(t) = \prod_{i=1}^m (e^t p + 1 - p)^{n_i} = (e^t p + 1 - p)^{\sum_{i=1}^m n_i} = \phi_Y(t),$$

where $Y \sim \text{BIN}(\sum_{i=1}^m n_i, p)$. It follows from (1.20) that $\sum_{i=1}^m X_i \sim \text{BIN}(\sum_{i=1}^m n_i, p)$. ◁

Convergence of a Sequence of Random Variables

Def'n 1.16

Let $(X_n)_{n=1}^\infty$ be a sequence of random variables and let X be a random variable.

- (a) We say $(X_n)_{n=1}^\infty$ **converges** to X **in distribution** if

$$\lim_{n \rightarrow \infty} \mathbb{P}\{X_n \leq x\} = \mathbb{P}\{X \leq x\}$$

for all $x \in \mathbb{R}$ at which F_X is continuous.

- (b) We say $(X_n)_{n=1}^\infty$ **converges** to X **in probability** if

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|X_n - X| > \varepsilon\} = 0$$

for every $\varepsilon > 0$.

(c) We say $(X_n)_{n=1}^{\infty}$ **converges** to X **almost surely (a.s.)** if

$$\mathbb{P} \left\{ \lim_{n \rightarrow \infty} X_n = X \right\} = 1.$$

(1.27)

Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables and let X be a random variable. Then

$$\begin{aligned} (X_n)_{n=1}^{\infty} \text{ converges to } X \text{ a.s.} &\implies (X_n)_{n=1}^{\infty} \text{ converges to } X \text{ in probability} \\ &\implies (X_n)_{n=1}^{\infty} \text{ converges to } X \text{ in distribution.} \end{aligned}$$

Theorem 1.6

Strong Law of Large Numbers (SLLN)

Let $(X_n)_{n=1}^{\infty}$ be a sequence of iid random variables with common expectation $\mu \in \mathbb{R}$. Then $(\bar{X}_n)_{n=1}^{\infty}$ converges to μ almost surely, where for every $n \in \mathbb{N}$,

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}.$$

This page intentionally left blank.

2.

Conditional Distributions

-
- 2.1 Jointly Discrete Case
 - 2.2 Jointly Continuous Case
 - 2.3 Conditioning
-

2.1 Jointly Discrete Case

(2.1) For convenience, we shall only consider *bivariate* case. Let X_1, X_2 be discrete random variables and let $x_2 \in \mathbb{R}$ throughout this section.

Def'n 2.1 Conditional PMF

If $p_{X_2}(x_2) > 0$, then we define the **conditional pmf** of X_1 given $X_2 = x_2$, denoted as $p_{X_1|X_2}(\cdot|x_2)$, is defined by

$$p_{X_1|X_2}(x_1|x_2) = \frac{p_{(X_1, X_2)}(x_1, x_2)}{p_{X_2}(x_2)}$$

for all $x_1 \in \mathbb{R}$. We denote the resulting distribution by $X_1|(X_2 = x_2)$.

(2.2) (a) We alternatively write $\mathbb{P}(X_1 = \cdot | X_2 = x_2)$ to denote $p_{X_1|X_2}(\cdot|x_2)$. Also note that

$$p_{X_1|X_2}(x_1|x_2) = \mathbb{P}(X_1 = x_1) = \frac{\mathbb{P}(X_1 = x_2, X_2 = x_2)}{\mathbb{P}(X_2 = x_2)} = \frac{p_{(X_1, X_2)}(x_1, x_2)}{p_{X_2}(x_2)}$$

(b) If X_1, X_2 are *independent*, then

$$p_{(X_1, X_2)}(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2)$$

for every $x_1, x_2 \in \mathbb{R}$, which means

$$p_{X_1|X_2}(x_1|x_2) = p_{X_1}(x_1)$$

for all $x_1, x_2 \in \mathbb{R}$ such that $p_{X_2}(x_2) > 0$.

Def'n 2.2 Conditional Expectation

If $p_{X_2}(x_2) > 0$, then we define the **conditional mean**, denoted as $\mathbb{E}(X_1|X_2 = x_2)$, of $X_1|(X_2 = x_2)$ by

$$\mathbb{E}(X_1|X_2 = x_2) = \sum_{x_1 \in \mathbb{R}: p_{X_1|X_2}(x_1|x_2) > 0} x_1 p_{X_1|X_2}(x_1|x_2).$$

Proposition 2.1

Let $w : \mathbb{R}^2 \rightarrow \mathbb{R}$. Then

$$\mathbb{E}(w(X_1, X_2) | X_2 = x_2) = \mathbb{E}(w(X_1, x_2) | X_2 = x_2).$$

Corollary 2.1.1

Given any $g, h : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E}(g(X_1)h(X_2) | X_2 = x_2) = \mathbb{E}(g(X_1)h(x_2) | X_2 = x_2).$$

Corollary 2.1.2

Let X_3 be a random variable and let $x_3 \in \mathbb{R}$ be such that $p_{X_3}(x_3) > 0$. Then

$$\mathbb{E}(X_1 + X_2 | X_3 = x_3) = \mathbb{E}(X_1 | X_3 = x_3) + \mathbb{E}(X_2 | X_3 = x_3).$$

Conditional Variance

Def'n 2.3

We define the **conditional variance** of $X_1|X_2 = x_2$, denoted as $\text{var}(X_1|X_2 = x_2)$, by

$$\text{var}(X_1|X_2 = x_2) = \mathbb{E}\left((X_1 - \mathbb{E}(X_1|X_2 = x_2))^2 | X_2 = x_2\right).$$

Proposition 2.2

We have

$$\text{var}(X_1|X_2 = x_2) = \mathbb{E}(X_1^2|X_2 = x_2) - \mathbb{E}(X_1|X_2 = x_2)^2.$$

(EX 2.3)

Suppose $X_1 \sim \text{BIN}(n_1, p)$, $X_2 \sim \text{BIN}(n_2, p)$ for some $n_1, n_2 \in \mathbb{N} \cup \{0\}$, $p \in (0, 1]$ are independent and let $m \in \mathbb{N} \cup \{0\}$. Find $p_{X_1|X_1+X_2}(\cdot | X_1 + X_2 = m)$.

Answer. We may assume $m \leq n_1 + n_2$, since otherwise $p_{X_1+X_2}(m) = 0$. Then observe that

$$\begin{aligned} p_{X_1|X_1+X_2}(x_1 | X_1 + X_2 = m) &= \frac{\mathbb{P}(X_1 = x_1, X_1 + X_2 = m)}{\mathbb{P}(X_1 + X_2 = m)} \\ &= \frac{\mathbb{P}(X_1 = x_1, X_2 = m - x_1)}{\mathbb{P}(X_1 + X_2 = m)} \\ &= \frac{\mathbb{P}(X_1 = x_1) \mathbb{P}(X_2 = m - x_1)}{\mathbb{P}(X_1 + X_2 = m)} && \text{since } X_1, X_2 \text{ are independent} \\ &= \frac{\binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \binom{n_2}{m-x_1} p^{m-x_1} (1-p)^{n_2-m+x_1}}{\binom{n_1+n_2}{m} p^m (1-p)^{n_1+n_2-m}} && \text{since } X_1+X_2 \sim \text{BIN}(n_1+n_2, p) \end{aligned}$$

for all $x_1 \in \{0, \dots, n_1\}$. But note that this is exactly the pmf of $\text{HG}(n_1 + n_2, n_1, m)$. That is,

$$X_1 | (X_1 + X_2 = m) \sim \text{HG}(n_1 + n_2, n_1, m).$$

◁

Here is an intuitive explanation of why $X_1 | (X_1 + X_2 = m) \sim \text{HG}(n_1 + n_2, n_1, m)$. Consider a sequence of $n_1 + n_2$ Bernoulli trials $(B_i)_{i=1}^{n_1+n_2}$, each with success probability p . We know exactly m of $B_1, \dots, B_{n_1+n_2}$ are successes, and we also know exactly n_1 of B_1, \dots, B_{n_1} are successes. But each B_i has success probability p , so we end up with a hypergeometric distribution. See (1.11).

(EX 2.4)

Let $X_1 \sim \text{POI}(\lambda_1), \dots, X_m \sim \text{POI}(\lambda_m)$ for some $\lambda_1, \dots, \lambda_m > 0$ be independent and let $Y = \sum_{i=1}^m X_i$. Find the conditional distribution of $X_j | (Y = n)$, where $j \in \{1, \dots, m\}$, $n \in \mathbb{N}$.

Answer. First note that $X_j, \sum_{i=1, i \neq j}^m X_i$ are independent, since X_1, \dots, X_m are independent. Fix $x_j \in \{0, \dots, n\}$. Then

$$\begin{aligned} p_{X_j|Y}(x_j | n) &= \frac{\mathbb{P}(X_j = x_j, Y = n)}{\mathbb{P}(Y = n)} \\ &= \frac{\mathbb{P}(X_j = x_j, \sum_{i=1}^m X_i = n)}{\mathbb{P}(Y = n)} \\ &= \frac{\mathbb{P}(X_j = x_j, \sum_{i=1, i \neq j}^m X_i = n - x_j)}{\mathbb{P}(Y = n)} \\ &= \frac{\mathbb{P}(X_j = x_j) \mathbb{P}(\sum_{i=1, i \neq j}^m X_i = n - x_j)}{\mathbb{P}(Y = n)}. \end{aligned}$$

since $X_j, \sum_{i=1, i \neq j}^m X_i$ are independent

But

$$Y \sim \text{POI}\left(\sum_{i=1}^m \lambda_i\right), \sum_{i=1, i \neq j}^m X_i \sim \text{POI}\left(\sum_{i=1, i \neq j}^m \lambda_i\right) \quad [2.1]$$

as sums of random variables, so

$$\begin{aligned} p_{X_j|Y}(x_j|n) &= \frac{\frac{e^{-\lambda_j} \lambda_j^{x_j}}{x_j!} e^{-\sum_{i=1, i \neq j}^m \lambda_i} \left(\sum_{i=1, i \neq j}^m \lambda_i\right)^{n-x_j}}{\frac{e^{\sum_{i=1}^m \lambda_i} \left(\sum_{i=1}^m \lambda_i\right)^n}{n!}}. && \text{by [2.1]} \\ &= \binom{n}{x_j} \frac{\lambda_j^{x_j} \left(\sum_{i=1, i \neq j}^m \lambda_i\right)^{n-x_j}}{\left(\sum_{i=1}^m \lambda_i\right)^n} \\ &= \binom{n}{x_j} \frac{\lambda_j^{x_j} (\lambda - \lambda_j)^{n-x_j}}{\lambda^n} && \text{by letting } \lambda = \sum_{i=1}^m \lambda_i \\ &= \binom{n}{x_j} \left(\frac{\lambda_j}{\lambda}\right)^{x_j} \left(\frac{\lambda - \lambda_j}{\lambda}\right)^{n-x_j} \\ &= \binom{n}{x_j} \left(\frac{\lambda_j}{\lambda}\right)^{x_j} \left(1 - \frac{\lambda_j}{\lambda}\right)^{n-x_j} \\ &= \binom{n}{x_j} p^{x_j} (1-p)^{n-x_j}. && \text{by letting } p = \frac{\lambda_j}{\lambda} \end{aligned}$$

Since $0 < \lambda_i \leq \lambda$, $p \in (0, 1]$, so it follows that

$$X_j|Y = n \sim \text{BIN}\left(n, \frac{\lambda_j}{\sum_{i=1}^m \lambda_i}\right). \quad \triangleleft$$

2.2 Jointly Continuous Case

(2.5) Let X, Y be jointly continuous random variables and let $y \in \mathbb{R}$ throughout this section.

Def'n 2.4 **Conditional PDF**
We define the **conditional pdf** of X given $Y = y$, denoted as $f_{X|Y}(\cdot|y)$, by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f(y)}$$

for all $x \in \mathbb{R}$.

(2.6) Given $a, b \in \mathbb{R}, a \leq b$, observe that

$$\mathbb{P}(a \leq X \leq b|Y = y) = \int_a^b f_{X|Y}(x|y) \, dx.$$

Conditional Expectation

Def'n 2.5 We define the **conditional expectation** of X given $Y = y$, denoted as $\mathbb{E}(X|Y = y)$, as

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx.$$

Proposition 2.3

Let $g : \mathbb{R} \rightarrow \mathbb{R}$. Then

$$\mathbb{E}(g(X)|Y = y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx.$$

Conditional Variance

Def'n 2.6 We define the **conditional variance** of X given $Y = y$, denoted as $\text{var}(X|Y = y)$, as

$$\text{var}(X|Y = y) = \mathbb{E}\left((X - \mathbb{E}(X|Y = y))^2 | Y = y\right).$$

Proposition 2.4

We have

$$\text{var}(X|Y = y) = \mathbb{E}(X^2|Y = y) - \mathbb{E}(X|Y = y)^2.$$

2.3 Conditioning

(2.7)

Let X, Y be random variables. Then we can define $v : \mathbb{R} \rightarrow \mathbb{R}$ by

$$v(y) = \mathbb{E}(X|Y = y)$$

for all $y \in \mathbb{R}$.

$\mathbb{E}(X|Y)$

Notation 2.7 Consider the setting of (2.7). We write $\mathbb{E}(X|Y)$ to denote $v(Y)$.

Since any real-valued function of a random variable is a random variable, so it makes sense to consider the expectation of $\mathbb{E}(X|Y)$:

$$\mathbb{E}(\mathbb{E}(X|Y)) = \begin{cases} \sum_{y \in \mathbb{R}: p_Y(y) > 0} \mathbb{E}(X|Y = y) p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}(X|Y = y) f_Y(y) dy & \text{if } Y \text{ is continuous} \end{cases}. \quad [2.2]$$

Theorem 2.5

Law of Total Expectation

Let X, Y be random variables. Then

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y)).$$

Proof. We shall consider the continuous case only — assume X, Y are jointly continuous. Recall from the definition of $\mathbb{E}(X|Y)$ that

$$\mathbb{E}(\mathbb{E}(X|Y)) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y = y) f_Y(y) dy.$$

But

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \\ &= \int_{-\infty}^{\infty} x \frac{f_{(X,Y)}(x,y)}{f_Y(y)} dx.\end{aligned}$$

It follows that

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X|Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \frac{f_{(X,Y)}(x,y)}{f_Y(y)} dx f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{(X,Y)}(x,y) dx dy \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{(X,Y)}(x,y) dy dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \mathbb{E}(X),\end{aligned}$$

as desired. ■

(2.8)

Suppose $X \sim \text{GEO}_t(p)$ where $p \in (0, 1]$. Calculate $\mathbb{E}(X)$, $\text{var}(X)$ using the law of total expectation.

Answer. Recall that X is the number of iid Bernoulli trials, each with success probability p , needed to obtain the first success. So let Y be the first trial. Then observe that

$$X|(Y=1) = 1$$

but

$$X|(Y=0) = X + 1.$$

By the law of total expectation,

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y)) = p_Y(0) \mathbb{E}(X|Y=0) + p_Y(1) \mathbb{E}(X|Y=1) \\ &= (1-p) \mathbb{E}(X+1) + p \mathbb{E}(1) = (1-p) + (1-p) \mathbb{E}(X) + p = 1 + (1-p) \mathbb{E}(X),\end{aligned}$$

so rearranging gives

$$\mathbb{E}(X) = \frac{1}{p}.$$

On the other hand,

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}(\mathbb{E}(X^2|Y)) = p_Y(0) \mathbb{E}(X^2|Y=0) + p_Y(1) \mathbb{E}(X^2|Y=1) \\ &= (1-p) \mathbb{E}(X^2 + 2X + 1) + p \mathbb{E}(1) = (1-p) \mathbb{E}(X^2) + 2(1-p) \mathbb{E}(X) + 1,\end{aligned}$$

so

$$\mathbb{E}(X^2) = \frac{2(1-p) \mathbb{E}(X) + 1}{p} = \frac{\frac{2-p}{p} + 1}{p} = \frac{2}{p^2} - \frac{1}{p} + \frac{1}{p} = \frac{2}{p^2}.$$

Thus

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1}{p^2}.$$

◁

Note that the obtained expectation and variance agree with the known results.

Notation 2.8 $\text{var}(X|Y)$
Let X, Y be random variables. Let $v : \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$v(y) = \text{var}(X|Y = y)$$

for all $y \in \mathbb{R}$. Then we write $\text{var}(X|Y)$ to denote $v(Y)$.

(2.9)

Similar to $\mathbb{E}(X|Y)$, $\text{var}(X|Y)$ is a random variable as a function, v , of a random variable, Y . The following is a *variance analogue* of the law of total probability.

Theorem 2.6

Conditional Variance Formula

Let X, Y be random variables. Then

$$\text{var}(X) = \mathbb{E}(\text{var}(X|Y)) + \text{var}(\mathbb{E}(X|Y)).$$

Proof. First note that, for any $y \in \mathbb{R}$,

$$\text{var}(X|Y = y) = \mathbb{E}(X^2|Y = y) - \mathbb{E}(X|Y = y)^2,$$

which means

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

On the other hand,

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

It follows from the law of total expectation that

$$\mathbb{E}(\text{var}(X|Y)) + \text{var}(\mathbb{E}(X|Y)) = \mathbb{E}(\mathbb{E}(X^2|Y)) - \mathbb{E}(\mathbb{E}(X|Y))^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \text{var}(X). \quad \blacksquare$$

(EX 2.10)

Random Sum

Let $(X_i)_{i=1}^\infty$ be an iid sequence of random variables with common mean $\mu \in \mathbb{R}$ and common variance $\sigma^2 \geq 0$ and let N be a nonnegative integer-valued random variable that is independent of X_1, \dots . Let

$$T = \sum_{i=1}^N X_i.$$

Find $\mathbb{E}(T)$, $\text{var}(T)$.

Answer. By the law of total probability,

$$\begin{aligned} \mathbb{E}(T) &= \mathbb{E}(\mathbb{E}(T|N)) = \mathbb{E}(\mathbb{E}(T|N = n) |_{n=N}) = \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^N X_i | N = n\right) |_{n=N}\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^n X_i | N = n\right) |_{n=N}\right) = \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^n X_i\right) |_{n=N}\right) = \mathbb{E}\left(\sum_{i=1}^N X_i\right) \\ &= \mathbb{E}(\mu N) = \mu \mathbb{E}(N). \end{aligned}$$

Moreover,

$$\text{var}(T|N = n) = \text{var}\left(\sum_{i=1}^N X_i | N = n\right) = \text{var}\left(\sum_{i=1}^n X_i | N = n\right) = \text{var}\left(\sum_{i=1}^n X_i\right) = n\sigma^2,$$

which means

$$\mathbb{E}(\text{var}(T|N)) = \mathbb{E}(N\sigma^2) = \sigma^2 \mathbb{E}(N).$$

On the other hand,

$$\text{var}(\mathbb{E}(T|N)) = \text{var}(\mu N) = \mu^2 \text{var}(N).$$

Thus

$$\text{var}(T) = \mathbb{E}(\text{var}(T|N)) + \text{var}(\mathbb{E}(T|N)) = \sigma^2 \mathbb{E}(N) + \mu^2 \text{var}(N)$$

by the conditional variance formula. ◁

(2.11)

Recall from [2.2] that, given any random variables X, Y ,

$$\mathbb{E}(\mathbb{E}(X|Y)) = \begin{cases} \sum_{y \in \mathbb{R}: p_Y(y) > 0} \mathbb{E}(X|Y=y) p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y) f_Y(y) dy & \text{if } Y \text{ is continuous} \end{cases}.$$

Now, suppose that A represents some event of interest and we desire to determine $\mathbb{P}(A)$. Define an *indicator random variable* X by

$$X = \begin{cases} 0 & \text{if } A^C \text{ occurs} \\ 1 & \text{if } A \text{ occurs} \end{cases}.$$

Clearly, $\mathbb{P}(X=1) = \mathbb{P}(A)$, $\mathbb{P}(X=0) = 1 - \mathbb{P}(A)$, so that $X \sim B(\mathbb{P}(A))$. Hence $\mathbb{E}(X) = \mathbb{P}(A)$ and

$$\begin{aligned} \mathbb{E}(X|Y=y) &= \sum_{x \in \{0,1\}} x \mathbb{P}(X=x|Y=y) \\ &= 0 \mathbb{P}(X=0|Y=y) + 1 \mathbb{P}(X=1|Y=y) \\ &= \mathbb{P}(X=1|Y=y) \\ &= \mathbb{P}(A|Y=y). \end{aligned}$$

for any random variable Y . Hence [2.2] becomes

$$\mathbb{P}(A) = \begin{cases} \sum_{y \in \mathbb{R}: p_Y(y) > 0} \mathbb{E}(A|Y=y) p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}(A|Y=y) f_Y(y) dy & \text{if } Y \text{ is continuous} \end{cases} \quad [2.3]$$

for all random variable Y .