Stochastic Processes 1

STAT 333

Fall 2021 (1219)

LaTeXer: *Cameron Roopnarine*

Instructors: *Steve Drekic, Mirabelle Huynh*

13th September 2021

Contents

Chapter 1

Review of Elementary Probability

Week 1
8th to 15th September

1.1 Fundamental Definition of a Probability Function

Probability Model: A probability model consists of 3 essential components: a *sample space*, a collection of *events*, and a *probability function (measure)*.

- **Sample Space**: For a random experiment in which all possible outcomes are known, the set of all possible outcomes is called the sample space (denoted by Ω).
- **Event**: Every subset A of a sample space Ω is an event.
- **Probability Function**: For each event A of Ω, $\mathbb{P}(A)$ is defined as the *probability of an event* A, satisfying 3 conditions:

  (i) $0 \leq \mathbb{P}(A) \leq 1$,

  (ii) $\mathbb{P}(\Omega) = 1$, or equivalently, $\mathbb{P}(\emptyset) = 0$, where $\emptyset$ is the *null event*,

  (iii) For $n \in \mathbb{Z}^+$ (in fact, $n = \infty$ as well), $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ if the sequence of events $\{A_i\}_{i=1}^n$ is *mutually exclusive* (i.e., $A_i \cap A_j = \emptyset \ \forall i \neq j$).

As a result of conditions (ii) and (iii), and noting that A^c is the complement of A, it follows that

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) \implies \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

1.2 Conditional Probability

Conditional Probability: The *conditional probability of event A given event B occurs* is defined as

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

provided that $\mathbb{P}(B) > 0$.

Remarks:

(1) When $B = \Omega$, $\mathbb{P}(A \mid \Omega) = \mathbb{P}(A \cap \Omega)/\,\mathbb{P}(\Omega) = \mathbb{P}(A)/1 = \mathbb{P}(A)$, as one would expect.

(2) Rewriting the above formula, $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B)$, which is often referred to as the basic "multiplication rule." For a sequence of events $\{A_i\}_{i=1}^{n}$, the generalized multiplication rule is given by

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Example 1.1. Suppose that we roll a fair six-sided die once (i.e., $\Omega = \{1, 2, 3, 4, 5, 6\}$). Let A denote the event of rolling a number less than 4 (i.e., $A = \{1, 2, 3\}$), and let B denote the event of rolling an odd number (i.e., $B = \{1, 3, 5\}$). Given that the roll is odd, what is the probability that number rolled is less than 4?

Solution: Since the die is fair, it immediately follows that $\mathbb{P}(A) = 3/6 = 1/2$ and $\mathbb{P}(B) = 3/6 = 1/2$. Moreover,

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(\{1, 2, 3\} \cap \{1, 3, 5\}) && (1.1)\\
&= \mathbb{P}(\{1, 3\}) && (1.2)\\
&= \frac{2}{6} && (1.3)\\
&= \frac{1}{3}. && (1.4)
\end{aligned}$$

Therefore,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

1.3 Independence of Events

Independence of Events: Two events A and B are *independent* if and only if (iff)

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

In general, if an experiment consists of a sequence of independent trials, and $A_1, A_2, \ldots, A_n$ are events such that A_i depends only on the i^{th} trial, then $A_1, A_2, \ldots, A_n$ are independent events and

$$\mathbb{P}(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

1.4 Law of Total Probability

Law of Total Probability: For $n \in \mathbb{Z}^+$ (and even $n = \infty$), suppose that $\Omega = \cup_{i=1}^{n} B_i$, where the sequence of events $\{B_i\}_{i=1}^n$ is mutually exclusive. Then,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) \\ &= \mathbb{P}(A \cap \{\cup_{i=1}^n B_i\}) \\ &= \mathbb{P}(\cup_{i=1}^n \{A \cap B_i\}) \\ &= \sum_{i=1}^n \mathbb{P}(A \cap B_i) \\ &= \sum_{i=1}^n \mathbb{P}(A \mid B_i)\, \mathbb{P}(B_i),\end{aligned}$$

where the second last equality follows from the fact that the sequence of events $\{A \cap B_i\}_{i=1}^n$ is also mutually exclusive.

1.5 Bayes' Formula

Bayes' Formula: Under the same assumptions as in the previous slide,

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_j)\, \mathbb{P}(B_j)}{\sum_{i=1}^n \mathbb{P}(A \mid B_i)\, \mathbb{P}(B_i)}.$$

1.6 Definition of a Random Variable

Definition: A *random variable* (rv) X is a real-valued function which maps a sample space Ω onto a state space $\mathcal{S} \subseteq \mathbb{R}$ (i.e., $X\colon \Omega \to \mathcal{S}$).

Discrete type: $\mathcal{S}$ consists of a finite or countable number of possible values. Important functions include:

$$\begin{aligned} p(a) &= \mathbb{P}(X = a) && \text{(pmf)}, \\ F(a) &= \mathbb{P}(X \le a) = \sum_{x \le a} p(x) && \text{(cdf)}, \\ \bar{F}(a) &= \mathbb{P}(X > a) = 1 - F(a) && \text{(tpf)}, \end{aligned}$$

where pmf stands for *probability mass function*, cdf stands for *cumulative distribution function*, and tpf stands for *tail probability function*.

Remark: If X takes on values in the set $\mathcal{S} = \{a_1, a_2, a_3, \dots\}$ where $a_1 < a_2 < a_3 < \cdots$ such that $p(a_i) > 0\ \forall i$, then we can recover the pmf from knowledge of the cdf via

$$\begin{aligned} p(a_1) &= F(a_1), \\ p(a_i) &= F(a_i) - F(a_{i-1}),\ i = 2, 3, 4, \dots \end{aligned}$$

1.7 Discrete Distributions

Special Discrete Distributions:

1. **Bernoulli**: If we consider a *Bernoulli trial*, which is a random trial with probability p of being a "success" (denoted by 1) and a probability $1-p$ of being a "failure" (denoted by 0), then X is *Bernoulli* (i.e., $X \sim \text{BERN}(p)$) with pmf

$$p(x) = p^x(1-p)^{1-x}, \; x = 0, 1.$$

2. **Binomial**: If X denotes the number of successes in $n \in \mathbb{Z}^+$ independent Bernoulli trials, each with probability p of being a success, then X is Binomial (i.e., $X \sim \text{BIN}(n,p)$) with pmf

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x}, \; x = 0, 1, \ldots, n,$$

where

$$\binom{n}{x} = \frac{n!}{(n-x)!x!} = \frac{(n)_x}{x!} = \frac{n(n-1)\cdots(n-x+1)}{x!}$$

is the number of distinct groups of x objects chosen from a set of n objects.

Remarks:

(1) A $\text{BIN}(1,p)$ distribution simplifies to become the $\text{BERN}(p)$ distribution.

(2) The binomial pmf is even defined for $n = 0$, in which case $p(x) = 1$ for $x = 0$. Such a distribution is said to be degenerate at 0.

(3) Note that $\binom{n}{x} = 0$ if $n, x \in \mathbb{N}$ with $n < x$.

3. **Negative Binomial**: If X denotes the number of Bernoulli trials (each with success probability p) required to observe $k \in \mathbb{Z}^+$ successes, then X is *Negative Binomial* (i.e., $X \sim \text{NB}_t(k,p)$) with pmf

$$p(x) = \binom{x-1}{k-1} p^k(1-p)^{x-k}, \; x = k, k+1, k+2, \ldots .$$

Remarks:

(1) In the above pmf, $\binom{x-1}{k-1}$ appears rather than $\binom{x}{k}$ since the final trial must always be a success.

(2) Sometimes, a negative binomial distribution is alternatively defined as the number of failures observed to achieve k successes. If Y denotes such a rv and $X \sim \text{NB}_t(k,p)$, then we clearly have the relationship $X = Y + k$, which immediately leads to the following pmf for Y:

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(X = y + k) = \binom{y+k-1}{k-1} p^k(1-p)^y, \; y = 0, 1, 2, \ldots .$$

To refer to this negative binomial distribution, we will write $Y \sim \text{NB}_f(k,p)$.

4. **Geometric**: If $X \sim \text{NB}_t(1,p)$, then X is *Geometric* (i.e., $X \sim \text{GEO}_t(p)$) with pmf

$$p(x) = p(1-p)^{x-1}, \; x = 1, 2, 3 \ldots .$$

In other words, the geometric distribution models the number of Bernoulli trials required to observe the first success.

Remark: Similarly, if $X \sim \text{NB}_f(1, p)$ then we obtain an alternative geometric distribution (denoted by $X \sim \text{GEO}_f(p)$) which models the number of failures observed prior to the first success.

5. **Discrete Uniform**: If X is equally likely to take on values in the (finite) set $\{a, a+1, \ldots, b\}$ where $a, b \in \mathbb{Z}$ with $a \leq b$, then X is *Discrete Uniform* (i.e., $X \sim \text{DU}(a, b)$) with pmf

$$p(x) = \frac{1}{b-a+1},\ x = a, a+1, \ldots, b.$$

6. **Hypergeometric**: If X denotes the number of success objects in n draws without replacement from a finite population of size N containing exactly r success objects, then X is *Hypergeometric* (i.e., $X \sim \text{HG}(N, r, n)$) with pmf

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},\ x = \max\{0, n-N+r\}, \ldots, \min\{n, r\}.$$

7. **Poisson**: A rv X is *Poisson* (i.e., $X \sim \text{POI}(\lambda)$) with parameter $\lambda > 0$ if its pmf is one of the form

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, 2, \ldots.$$

Remark: The pmf is even defined for $\lambda = 0$ (if we use the standard convention that $0^0 = 1$), in which case $p(x) = 1$ for $x = 0$ (i.e., X is degenerate at 0).

Example 1.2. Show that when n is large and p is small, the $\text{BIN}(n, p)$ distribution may be approximated by a $\text{POI}(\lambda)$ distribution where $\lambda = np$.

Solution: Recall $e^z = \lim\limits_{n\to\infty} (1 + z/n)^n$, $z \in \mathbb{R}$. Letting $X \sim \text{BIN}(n, p)$, we have

$$\begin{aligned}
\mathbb{P}(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \\
&= \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n}\frac{\lambda^x}{x!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^x} \\
&\simeq (1)(1)\cdots(1)\frac{\lambda^x}{x!}\frac{e^{-\lambda}}{1} && \text{when } n \text{ is large} \\
&= \frac{e^{-\lambda}\lambda^x}{x!}
\end{aligned}$$

1.8 Continuous Random Variables

Continuous type: A rv X takes on a continuum of possible values (which is uncountable) with cdf

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y,$$

where $f(x)$ denotes the *probability density function* (pdf) of X, which is a non-negative real-valued function that satisfies

$$\mathbb{P}(X \in B) = \int_{x \in B} f(x)\,\mathrm{d}x,$$

where B is the set of real numbers (e.g., an interval).

Remarks:

(1) If $F(x)$ (or the tpf $\bar{F}(x) = 1 - F(x)$) is known, we can recover the pdf using the relation

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) = F'(x) = -\bar{F}'(x),$$

which holds by the *Fundamental Theorem of Calculus.*

(2) When working with pdfs in general, it is usually not necessary to be precise about specifying whether a range of numbers includes the endpoints. This is quite different from the situation we encounter with discrete rvs. Throughout this course, however, we will adopt the convention of **not including** the endpoints when specifying the range of values for pdfs.

1.9 Continuous Distributions

Special Continuous Distributions:

1. **Uniform**: A rv X is *Uniform* on the real interval (a, b) (i.e., $X \sim \mathrm{U}(a, b)$) if it has pdf

$$f(x) = \frac{1}{b-a},\ a < x < b,$$

where $a, b \in \mathbb{R}$ with $a < b$.

Remark: The choice of name is because X takes on values in (a, b) with all subintervals of a fixed length being equally likely.

2. **Beta**: A rv X is *Beta* with parameters $m \in \mathbb{Z}^+$ and $n \in \mathbb{Z}^+$ (i.e., $X \sim \mathrm{Beta}(m, n)$) if it has pdf

$$f(x) = \frac{(m+n-1)!}{(m-1)!(n-1)!}x^{m-1}(1-x)^{n-1},\ 0 < x < 1.$$

Remark: A Beta$(1, 1)$ distribution simplifies to become the U$(0, 1)$ distribution.

3. **Erlang**: A rv X is *Erlang* with parameters $n \in \mathbb{Z}^+$ and $\lambda > 0$ (i.e., $X \sim \mathrm{Erlang}(n, \lambda)$) if it has

pdf

$$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!},\ x > 0.$$

Remark: The Erlang(n, λ) distribution is actually a special case of the more general Gamma distribution in which n is extended to be any positive real number.

4. **Exponential**: A rv X is *Exponential* with parameter $\lambda > 0$ (i.e., $X \sim \text{EXP}(\lambda)$) if it has pdf

$$f(x) = \lambda e^{-\lambda x},\ x > 0.$$

Remark: An Erlang$(1, \lambda)$ distribution actually simplifies to become the EXP(λ) distribution.

1.10 Expectation

Expectation: If $g(\,\cdot\,)$ is an arbitrary real-valued function, then

$$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x)p(x) & \text{, if } X \text{ is a discrete rv,} \\ \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x & \text{, if } X \text{ is a continuous rv.} \end{cases}$$

Special choices of $g(\,\cdot\,)$:

1. $g(X) = X^n$, $n \in \mathbb{N} \implies \mathbb{E}[g(X)] = \mathbb{E}[X^n]$ is the n^{th} moment of X. In general, moments serve to describe the shape of a distribution. If $n = 0$, then $\mathbb{E}[X^0] = 1$. If $n = 1$, then $\mathbb{E}[X] = \mu_X$ is the *mean* of X.
2. $g(X) = (X - \mathbb{E}[X])^2 \implies \mathbb{E}[g(X)] = \mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big]$ is the *variance* of X. Note that

$$\text{Var}(X) = \sigma_X^2 = \mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big] = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

   or equivalently

$$\sigma_X^2 = \mathbb{E}[X - (X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2.$$

   Related to this quantity, the *standard deviation* of X is $\sqrt{\text{Var}(X)} = \sigma_X$.
3. $g(X) = aX + b$, $a, b \in \mathbb{R}$ (i.e., $g(X)$ is a linear function of X). Note that

$$\begin{aligned} \mu_{aX+b} &= \mathbb{E}[aX + b] = a\mu_X + b, \\ \sigma_{aX+b}^2 &= \text{Var}(aX + b) = a^2\sigma_X^2, \\ \sigma_{aX+b} &= \sqrt{\text{Var}(aX + b)} = |a|\sigma_X. \end{aligned}$$

1.11 Moment Generating Function

4. $g(X) = e^{tX}$, $t \in \mathbb{R} \implies \mathbb{E}[g(X)] = \mathbb{E}[e^{tX}]$ is the *moment generating function* (mgf) of X. This

quantity is a function of t and is denoted by

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

First, $\phi_X(0) = \mathbb{E}[e^{0X}] = \mathbb{E}[1] = 1$. Moreover, making use of the linearity property of the expected value operator, note that

$$\begin{aligned}
\phi_X(t) &= \mathbb{E}[e^{tX}] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right] \\
&= \mathbb{E}\left[\frac{t^0X^0}{0!} + \frac{t^1X^1}{1!} + \frac{t^2X^2}{2!} + \cdots + \frac{t^nX^n}{n!} + \cdots\right] \\
&= \mathbb{E}[X^0]\frac{t^0}{0!} + \mathbb{E}[X]\frac{t^1}{1!} + \mathbb{E}[X^2]\frac{t^2}{2!} + \cdots + \mathbb{E}[X^n]\frac{t^n}{n!} + \cdots,
\end{aligned}$$

implying that the n^{th} moment of X is simply the coefficient of $t^n/n!$ in the above series expansion.

We have: $\phi_X(t) = \mathbb{E}[t^X] = \mathbb{E}[X^0]\frac{t^0}{0!} + \mathbb{E}[X]\frac{t^1}{1!} + \mathbb{E}[X^2]\frac{t^2}{2!} + \cdots + \mathbb{E}[X^n]\frac{t^n}{n!} + \cdots$.

Remarks:

(1) Given the mgf of X, we can extract its n^{th} moment via

$$\mathbb{E}[X^n] = \phi_X^{(n)}(0) = \frac{\mathrm{d}^n}{\mathrm{d}t^n}\phi_X(t)\bigg|_{t=0} = \lim_{t\to 0}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\phi_X(t),\ n \in \mathbb{N}.$$

Note that the 0^{th} derivative of a function is simply the function itself.

(2) A mgf uniquely characterizes the probability distribution of a rv (i.e., there exists a one-to-one correspondence between the mgf and the pmf/pdf of a rv). In other words, if two rvs X and Y have the same mgf, then they must have the same probability distribution (which we denote by $X \sim Y$). Thus, by finding the mgf of a rv, one has indeed determined its probability distribution.

Example 1.3. Suppose that $X \sim \text{BIN}(n, p)$. Find the mgf of X and use it to find $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution: Recall the binomial series formula

$$(a+b)^m = \sum_{x=0}^{m} \binom{m}{x} a^x b^{m-x},\ a, b \in \mathbb{R},\ m \in \mathbb{N}.$$

Using this formula, we obtain

$$\begin{aligned}
\phi_X(t) &= \mathbb{E}[e^{tX}] \\
&= \sum_{x=0}^{n} e^{tx}\binom{n}{x} p^x(1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x(1-p)^{n-x} \\
&= (pe^t + 1 - p)^n,\ t \in \mathbb{R}.
\end{aligned}$$

Then,

$$\phi_X'(t) = n(pe^t + 1 - p)^{n-1}pe^t \quad \text{and} \quad Q_X''(t) = n(pe^t + 1 - p)^{n-1}pe^t + npe^t(n-1)(pe^t + 1 - p)^{n-2}pe^t.$$

Thus,

$$\mathbb{E}[X] = \phi_X'(0) = n(pe^0 + 1 - p)^{n-1}pe^0 = np,$$
$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \phi_X''(0) - n^2p^2 = np + np(n-1)p - n^2p^2 = np.$$

1.12 Joint Distributions

Joint Distributions: The following results are presented for the bivariate case mostly, but these ideas extend naturally to an arbitrary number of rvs.

Definition: The *joint cdf* of X and Y is

$$\begin{aligned} F(a,b) &= \mathbb{P}(X \le a, Y \le b) \\ &= \mathbb{P}(\{X \le a\} \cap \{Y \le b\}),\ a, b \in \mathbb{R}. \end{aligned}$$

Remark: If the joint cdf is known, then we can recover their marginal counterparts as follows:

$$F_X(a) = \mathbb{P}(X \le a) = F(a, \infty) = \lim_{b\to\infty} F(a,b),$$
$$F_Y(a) = \mathbb{P}(Y \le b) = F(\infty, b) = \lim_{a\to\infty} F(a,b).$$

Jointly Discrete Case:
Joint pmf:

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

Marginals:

$$p_X(x) = \mathbb{P}(X = x) = \sum_y p(x,y)$$
$$p_Y(y) = \mathbb{P}(Y = y) = \sum_x p(x,y)$$

Multinomial Distribution: Consider an experiment which is repeated $n \in \mathbb{Z}^+$ times, with one of $k \ge 2$ distinct outcomes possible each time. Let $p_1, p_2, \dots, p_k$ denote the probabilities of the k types of outcomes (with $\sum_{i=1}^k p_i = 1$). If X_i, $i = 1, 2, \dots, k$, counts the number of type-i outcomes to occur, then $(X_1, X_2, \dots, X_k)$ is *Multinomial* (i.e., $(X_1, X_2, \dots, X_k) \sim \text{MN}(n, p_1, p_2, \dots, p_k)$) with joint pmf

$$p(x_1, x_2, \dots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k},\ x_i = 0, 1, \dots, n\ \forall i \text{ and } \sum_{i=1}^k x_i = n$$

Remark: A $\text{MN}(n, p_1, 1 - p_1)$ distribution simplifies to become the $\text{BIN}(n, p_1)$ distribution.

Jointly Continuous Case:
Joint pdf: The joint pdf $f(x,y)$ is a non-negative real-valued function which enables one to calculate probabilities of the form

$$\mathbb{P}(X \in A, Y \in B) = \int_B \int_A f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_A \int_B f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

where A and B are sets of real numbers (e.g., intervals). As a result,

$$F(a,b) = \int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{a}\int_{-\infty}^{b} f(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

Marginals:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x \end{aligned}$$

Jointly Continuous Case:
Important Relationship:

$$f(x,y) = \frac{\partial^2}{\partial x\,\partial y} F(x,y)$$

Transformations: Let (X,Y) be jointly continuous with joint pdf $f(x,y0)$ and region of support $\mathcal{S}(X,Y)$. Suppose that the rvs V and W are given by $V = b_1(X,Y)$ and $W = b_2(X,Y)$, where the functions $v = b_1(x,y)$ and $w = b_2(x,y)$ defined a one-to-one transformation that maps the set $\mathcal{S}(X,Y)$ onto the set $\mathcal{S}(V,W)$. If x and y are expressed in terms of v and w (i.e., $x = h_1(v,w)$ and $y = h_2(v,w)$), then the joint pdf of V and W is given by

$$g(v,w) = \begin{cases} f(h_1(v,w), h_2(v,w))|J| & \text{, if } (v,w) \in \mathcal{S}(V,W), \\ 0 & \text{, elsewhere,} \end{cases}$$

where J is the *Jacobian* of the transformation given by

$$J = \frac{\partial x}{\partial v}\frac{\partial y}{\partial w} - \frac{\partial x}{\partial w}\frac{\partial y}{\partial v}.$$

1.13 Expectation

Expectation: If $g(\,\cdot\,,\,\cdot\,)$ denotes an arbitrary real-valued function, then

$$\mathbb{E}[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y) & \text{, if } X \text{ and } Y \text{ are jointly discrete,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,\mathrm{d}y\,\mathrm{d}x & \text{, if } X \text{ and } Y \text{ are jointly continuous.} \end{cases}$$

Remark: The order of summation/integration is irrelevant and can be interchanged.

Special choices of $g(\,\cdot\,)$:

1. $g(X,Y) = (X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) \implies \mathbb{E}[g(X,Y)] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ is the *covariance* of X and Y. Note that

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

and $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$.

2. $g(X,Y) = aX + bY$, $a, b \in \mathbb{R}$ (i.e., $g(X,Y)$ is a linear combination of X and Y). Note that:

$$\begin{aligned} \mathbb{E}[aX+bY] &= a\,\mathbb{E}[X] + b\,\mathbb{E}[Y], \\ \operatorname{Var}(aX+bY) &= a^2\operatorname{Var}(X) + b^2\operatorname{Var}(Y) + 2ab\operatorname{Cov}(X,Y). \end{aligned}$$

3. $g(X,Y) = e^{sX+tY}$, $s, t \in \mathbb{R} \implies \mathbb{E}[g(X,Y)] = \mathbb{E}[e^{sX+tY}]$ is the joint mgf of X and Y. A joint mgf (denoted by $\phi(s,t)$) also uniquely characterizes a joint probability distribution and can be used to calculate joint moments of X and Y via the formula

$$\mathbb{E}[X^mY^n] = \phi^{(m,n)}(0,0) = \left(\frac{\partial^{m+n}}{\partial s^m\,\partial t^n}\phi(s,t)\right)_{s=0,t=0} = \lim_{s\to 0, t\to 0}\frac{\partial^{m+n}}{\partial s^m\,\partial t^n}\phi(s,t),\ m, n \in \mathbb{N}$$

1.14 Independence of Random Variables

Formal Definition: If X and Y are *independent* rvs if

$$\begin{aligned} F(a,b) &= \mathbb{P}(X \le a, Y \le b) \\ &= \mathbb{P}(X \le a)\,\mathbb{P}(Y \le b) \\ &= F_X(a)F_Y(b)\ \forall a, b \in \mathbb{R}. \end{aligned}$$

Equivalently, independence exists iff $p(x,y) = p_X(x)p_Y(y)$ (in the jointly discrete case) or $f(x,y) = f_X(x)f_Y(y)$ (in the jointly continuous case) $\forall x, y \in \mathbb{R}$.

Important Property: For arbitrary real-valued functions $g(\,\cdot\,)$ and $h(\,\cdot\,)$, if X and Y are independent, then

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)].$$

Remark: As a consequence of this property, $\operatorname{Cov}(X,Y) = 0$ if X and Y are independent, implying that $\operatorname{Var}(aX+bY) = a^2\operatorname{Var}(X) + b^2\operatorname{Var}(Y)$. However, if $\operatorname{Cov}(X,Y) = 0$, we cannot conclude that X and Y are independent (we can only say that X and Y are *uncorrelated*).

Example 1.4. Suppose that X and Y have joint pmf (and corresponding marginals) of the form

		y		
	$p(x,y)$	0	1	$p_X(x)$
x	0	0.2	0	0.2
	1	0	0.6	0.6
	2	0.2	0	0.2
	$p_Y(y)$	0.4	0.6	1

Show that $\operatorname{Cov}(X,Y) = 0$ holds, but X and Y are not independent.

Solution: Recall that $\operatorname{Cov}(X,X) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. Note that

$$\begin{aligned} \mathbb{E}[XY] &= \sum_x\sum_y xyp(x,y) \\ &= (0)(0)(0.2) + (0)(1)(0) + (1)(0)(0) + (1)(1)(0.6) + (2)(0)(0.2) + (2)(1)(0) \\ &= 0.6, \end{aligned}$$

$$\begin{aligned}\mathbb{E}[X] &= \sum_x x p_X(x) = (0)(0.2) + (1)(0.6) + (2)(0.2) = 1,\\ \mathbb{E}[Y] &= \sum_y y p_Y(y) = (0)(0.4) + (1)(0.6) = 0.6.\end{aligned}$$

Thus,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0.6 - (1)(0.6) = 0.$$

However, from the given table, it is clear that $p(2,0) = 0.2 \neq 0.08 = (0.2)(0.4) = p_X(2)p_Y(0)$. Thus, we conclude that while $\mathrm{Cov}(X,Y) = 0$, X and Y are not independent.

Theorem 1.1. If $X_1, X_2, \dots, X_n$ are independent rvs where $\phi_{X_i}(t)$ is the mgf of X_i, $i = 1, 2, \dots, n$, then $T = \sum_{i=1}^n X_i$ has mgf $\phi_T(t) = \prod_{i=1}^n \phi_{X_i}(t)$.

Proof: Note that the mgf of T given by

$$\begin{aligned}\phi_T(t) &= \mathbb{E}[e^{tT}]\\ &= \mathbb{E}[e^{t(X_1+X_2+\dots+X_n)}]\\ &= \mathbb{E}[e^{tX_1}e^{tX_2}\dots e^{tX_n}]\\ &= \mathbb{E}[e^{tX_1}]\,\mathbb{E}[e^{tX_2}]\cdots\mathbb{E}[e^{tX_n}] && \text{by independence of } \{X_i\}_{i=1}^n\\ &= \phi_{X_1}(t)\phi_{X_2}(t)\cdots\phi_{X_n}(t)\\ &= \prod_{i=1}^n \phi_{X_i}(t).\end{aligned}$$

Remarks:

(1) Simply put, Theorem 1.1 states that the mgf of a sum of independent rvs is just the product of their individual mgfs.

(2) As a special case of the above result, note that $\phi_T(t) = \phi_{X_1}(t)^n$ if $X_1, X_2, \dots, X_n$ is an independent and identically distributed (iid) sequence of rvs.

Example 1.5. Let $X_1, X_2, \dots, X_m$ be an independent sequence of rvs where $X_i \sim \mathrm{BIN}(n_i, p)$, $i = 1, 2, \dots, m$. Find the distribution of $T = \sum_{i=1}^m X_i$.

Solution: Looking at the mgf of T, note that

$$\begin{aligned}\phi_T(t) &= \prod_{i=1}^m \phi_{X_i}(t) && \text{by Theorem 1.1}\\ &= \prod_{i=1}^m (pe^t + 1 - p)^{n_i} && \text{using the result of Example of 1.3}\\ &= (pe^t + 1 - p)^{\sum_{i=1}^m n_i},\ t \in \mathbb{R}.\end{aligned}$$

By the mgf uniqueness property we recognize that $T = \sum_{i=1}^m X_i \sim \mathrm{BIN}(\sum_{i=1}^m n_i, p)$.
Remark: As a special case of the above example, if $X_1, X_2, \dots, X_m$ are iid $\mathrm{BERN}(p)$ rvs, then $T = \sum_{i=1}^m X_i \sim \mathrm{BIN}(m, p)$.

1.15 Convergence of Random Variables

Modes of Convergence: If X_n, $n \in \mathbb{Z}^+$, and X are rvs, then

1. $X_n \to X$ *in distribution* iff

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x), \ \forall x \in \mathbb{R} \text{ at which } \mathbb{P}(X \le x) \text{ is continuous},$$

2. $X_n \to X$ *in probability*, iff $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0,$$

3. $X_n \to X$ *almost surely (a.s.)* iff

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1.$$

Remarks:

(1) In probability theory, an event is said to happen a.s. if it happens with probability 1.

(2) The following implications hold true in general:

$$X_n \to X \text{ a.s.} \implies X_n \to X \text{ in probability} \implies X_n \to X \text{ in distribution.}$$

1.16 Strong Law of Large Numbers

Strong Law of Large Numbers (SLLN): If $X_1, X_2, \dots, X_n$ is an iid sequence of rvs with common mean μ and $\mathbb{E}[|X_1|] < \infty$, then

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \text{ a.s. as } n \to \infty.$$

Remark: The SLLN is one of the most important results in probability and statistics, indicating that the sample mean will, with probability 1, converge to the true mean of the underlying distribution as the sample size approaches infinity. In other words, if the same experiment or study is repeated independently many times, the average of the results of the trials must be close to the mean. The result gets closer to the mean as the number of trials is increased.