EE 550: Multiple Access via Renewal Theory

Michael J. Neely
https://viterbi-web.usc.edu/~mjneely/

An updated version of these notes is on the following clickable link:
https://ee.usc.edu/stochastic-nets/docs/multi-access-renewal-theory.pdf

I. Introduction

These notes discuss multiple access techniques for data networks. This is part of the course material for the graduate course EE 550 at USC. First, the notes describe a result of *renewal theory* that follows directly from the law of large numbers. Next, the renewal theory result is applied to calculate throughput and energy expenditure in multiple access scenarios. A discussion of Robbins-Monro iterations for dynamically adjusting a transmit probability for basic multiple access scenarios is also discussed. The appendices of these notes contain details that may be of independent interest but are not required for the course.

A. Some references

Additional material on probability, Markov chains, and renewal theory is in [1][2]. A stabilized version of slotted Aloha is investigated in [3] for the case of an infinite number of users (all with one packet) when the number of users $N[t]$ is unknown. There, an estimate $\hat{N}[t]$ is used. The estimate is dynamically adjusted at the end of every slot based on idle/success/collision feedback. Multi-packet reception techniques using ZigZag and SigSag decoding are in [4][5], see also MAC protocols in [6][7]. Other "Coded-MAC" strategies that yield provably efficient random access schemes for multi-channel systems are in [8][9][10][11][12][13][14][15][16].

II. Probability review

A *probability space* is defined by a triplet $(\Omega, \mathcal{F}, P)$ where: Ω is a nonempty set called the *sample space*; $\mathcal{F}$ is a collection of subsets of Ω that satisfies the three properties of a *sigma algebra on* Ω (see Appendix A); $P : \mathcal{F} \to [0,1]$ is a function that satisfies the three axioms of probability (see Appendix A). The function P is called a *probability measure*.

If $A \subseteq \Omega$ then $A^c = \{\omega \in \Omega : \omega \notin A\}$. The empty set is the subset of Ω with no elements and is denoted ϕ. It holds that $\phi \subseteq \Omega$; $\Omega^c = \phi$; $\phi^c = \Omega$. Since $\mathcal{F}$ has the three properties of a sigma algebra on Ω, it can be shown that (see Appendix A):

- $\Omega \in \mathcal{F}$.
- $\phi \in \mathcal{F}$.
- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
- The union of a finite or countably infinite number of sets in $\mathcal{F}$ is again a set in $\mathcal{F}$.
- The intersection of a finite or countably infinite number of sets in $\mathcal{F}$ is again a set in $\mathcal{F}$.

Elements of Ω are called *outcomes*. Thus, $\omega \in \Omega$ if and only if ω is an outcome. Elements of $\mathcal{F}$ are called *events*. Thus, a set A is an event if and only if $A \in \mathcal{F}$. Since all elements of $\mathcal{F}$ are subsets of Ω, it holds that

$$(A \in \mathcal{F}) \implies (A \subseteq \Omega)$$

The reverse implication is only true in the special case when $\mathcal{F}$ is the set of all subsets of Ω. All events $A \in \mathcal{F}$ have probabilities $P[A]$. Subsets of Ω that are not in $\mathcal{F}$ are not events, do not have probabilities, and are called *nonmeasurable sets*.

A. Random variables

Fix a probability space $(\Omega, \mathcal{F}, P)$. A *random variable* is a function $X : \Omega \to \mathbb{R}$ that satisfies the following *measurability property*:[1]

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R} \tag{1}$$

The measurability property (1) automatically holds in the special case when $\mathcal{F}$ is the set of all subsets of Ω, and so every function $X : \Omega \to \mathbb{R}$ is a random variable in this special case.

[1] Fix $X : \Omega \to \mathbb{R}$. It can be shown that X satisfies (1) if and only if $\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ for all sets $B \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is the standard Borel sigma algebra on $\mathbb{R}$, defined as the smallest sigma algebra on $\mathbb{R}$ that contains all intervals of the type $(-\infty, x]$ for $x \in \mathbb{R}$. The sigma algebra $\mathcal{B}(\mathbb{R})$ is known to contain all finite or countably infinite subsets of $\mathbb{R}$; all open subsets of $\mathbb{R}$; all closed subsets of $\mathbb{R}$; all subsets of $\mathbb{R}$ that can be obtained by a countable procedure of complements, unions, or intersections of sets in $\mathcal{B}(\mathbb{R})$. Subsets of $\mathbb{R}$ that are not in $\mathcal{B}(\mathbb{R})$ are called *non-Borel sets*. Non-Borel sets typically do not arise in practice but can be proven to exist using the *axiom of choice*.

It is convenient to denote the set $\{\omega \in \Omega : X(\omega) \leq x\}$ with the compressed notation $\{X \leq x\}$. If X is a random variable, the measurability property implies that for each $x \in \mathbb{R}$, the set $\{X \leq x\}$ is a valid event, hence it has a probability. The *cumulative distribution function (CDF)* F_X is defined by collecting these probabilities into a function. Indeed, define $F_X : \mathbb{R} \rightarrow [0, 1]$ by

$$F_X(x) = P[X \leq x] \quad \forall x \in \mathbb{R}$$

where $P[X \leq x]$ is compressed notation for $P[\{X \leq x\}]$, that is

$$P[X \leq x] = P[\{X \leq x\}] = P\left[\{\omega \in \Omega : X(\omega) \leq x\}\right]$$

A probability experiment has one probability space $(\Omega, \mathcal{F}, P)$ and may involve many random variables $Y, Z, \{X_i\}_{i=1}^{\infty}$ *on that same probability space*, so they are all functions from Ω to $\mathbb{R}$:

$$\begin{aligned} &Y : \Omega \rightarrow \mathbb{R} \\ &Z : \Omega \rightarrow \mathbb{R} \\ &X_i : \Omega \rightarrow \mathbb{R} \quad \forall i \in \{1, 2, 3, \ldots\} \end{aligned}$$

and their measurability property is with respect to the same sigma algebra $\mathcal{F}$. The particular outcome $\omega \in \Omega$ determines the particular numerical values of all random variables on the space:

$$Y(\omega), Z(\omega), X_1(\omega), X_2(\omega), X_3(\omega), X_4(\omega)$$

and so on. Unless otherwise stated, when multiple random variables are considered together, it is implicitly assumed they are associated with the same probability experiment and hence are on the same probability space.[2]

If Y and Z are random variables the measurability property (1) ensures

$$\{Y \leq 4.5\} \in \mathcal{F}, \quad \{Z \leq -2\} \in \mathcal{F}$$

Since the intersection of two sets in $\mathcal{F}$ is again a set in $\mathcal{F}$, we conclude

$$\{Y \leq 4.5\} \cap \{Z \leq -2\} \in \mathcal{F}$$

Note that

$$\{Y \leq 4.5\} \cap \{Z \leq -2\} = \{\omega \in \Omega : Y(\omega) \leq 4.5 \text{ and } Z(\omega) \leq -2\}$$

It follows that $\{Y \leq 4.5\} \cap \{Z \leq -2\}$ is an event, so it has a valid probability $P[\{Y \leq 4.5\} \cap \{Z \leq -2\}]$. For simplicity of notation we often remove the braces and replace the intersection symbol with a comma (which means "and"):

$$P[Y \leq 4.5, Z \leq -2] = P[\{Y \leq 4.5\} \cap \{Z \leq -2\}] = P[\{\omega \in \Omega : Y(\omega) \leq 4.5 \text{ and } Z(\omega) \leq -2\}]$$

Similarly, if $\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables then for each $i \in \{1, 2, 3, \ldots\}$ and for each $x_i \in \mathbb{R}$, the measurability property ensures

$$\{X_i \leq x_i\} \in \mathcal{F}$$

The intersection of a finite number of events is again an event, so for any positive integer n we have

$$\cap_{i=1}^{n} \{X_i \leq x_i\} \in \mathcal{F}$$

In particular, $\cap_{i=1}^{n} \{X_i \leq x_i\}$ is an event and so it has a probability. The *joint cumulative distribution function (joint CDF)* $F_{X_1,\ldots,X_n}$ is defined by collecting these probabilities into a function. Define $F_{X_1,\ldots,X_n} : \mathbb{R}^n \rightarrow [0, 1]$ by

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P[\cap_{i=1}^{n} \{X_i \leq x_i\}] \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

B. Independence of random variables

Fix $(\Omega, \mathcal{F}, P)$ as a probability space. Fix n as a positive integer. Random variables $X_1, X_2, \ldots, X_n$ are defined as *mutually independent* if

$$P[\cap_{i=1}^{n} \{X_i \leq x_i\}] = \prod_{i=1}^{n} P[X_i \leq x_i] \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

This is equivalent to the joint CDF factoring into a product of marginal CDFs:

$$F_{X_1,\ldots X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n) \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

Random variables in an infinite sequence $\{X_i\}_{i=1}^{\infty}$ are defined to be *mutually independent* if $X_1, \ldots, X_n$ are mutually independent for each positive integer n.

[2]A single probability experiment may involve multiple steps. Typically, a single homework problem involves a single probability experiment. For example, if a homework problem involves randomly throwing twelve darts, then random variables X, Y, Z that appear in the problem are understood to be associated with that probability experiment. If another homework problem involves picking apples from a bin, the letters X, Y, Z can be reused to have a different meaning for that different probability experiment. It does not make sense to talk about the probability that "Y is less than or equal to 4.5 and Z is less than or equal to -2" unless the random variables Y and Z are part of the same probability experiment (meaning they are on the same probability space).

C. Operations on measurable functions

It can be shown that basic operations on random variables preserve the measurability property. Specifically, fix $(\Omega, \mathcal{F}, P)$ as a probability space. The sum or product of a finite number of random variables on this space produces another random variable on this space. A linear combination of a finite number of random variables on this space is another random variable on this space. For example, if $\{X_i\}_{i=1}^{\infty}$ is a sequence of random variables, then for each positive integer n we can define another random variable $M_n : \Omega \to \mathbb{R}$ by

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

meaning that

$$M_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \quad \forall \omega \in \Omega$$

More generally, it is known that continuous functions preserve measurability. Indeed, fix n as a positive integer and let $h : \mathbb{R}^n \to \mathbb{R}$ be any continuous function. Let $X_1, \ldots, X_n$ be random variables. Define $Y = h(X_1, \ldots, X_n)$, so that

$$Y(\omega) = h(X_1(\omega), \ldots, X_n(\omega)) \quad \forall \omega \in \Omega$$

It can be shown that Y satisfies the measurability property and hence Y is a random variable.[3]

D. Intuitive interpretation

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ be random variables. Let A be an event. An event is a subset of Ω that is in the collection $\mathcal{F}$. We often say things such as "if event A is true" or "if event A is false." How can a subset be "true" or "false"? It cannot. When we say "event A is true" we are imagining a situation where nature "randomly" selects an outcome $\omega \in \Omega$ according to the probability measure P. With this thought experiment, "event A is true" has the interpretation "the randomly selected outcome ω is in the set A." Similarly "event $\{X \leq 4.5 - Y\}$ is true" has the interpretation "the randomly selected outcome ω satisfies $X(\omega) \leq 4.5 - Y(\omega)$." Probability theory does not rely on such thought experiments. In fact, probability theory never defines the concept of "randomly choosing an outcome." However, such thought experiments are important for using probability theory to understand real world scenarios such as flipping coins or rolling dice, where the probability measure P is thoughtfully constructed to assign values to each event A according to our intuitive understanding of the "likelihood" of that event. The three axioms of probability (in Appendix A) also correspond to this intuition.

E. Summarizing exercises 1 (multiple choice)

1) In a probability triplet $(\Omega, \mathcal{F}, P)$, the set Ω is the
   a) set of probabilities
   b) set of events
   c) sample space
   d) set of expectations
2) In a probability triplet $(\Omega, \mathcal{F}, P)$, the collection $\mathcal{F}$ is the
   a) set of outcomes
   b) set of events
   c) sample space
   d) set of expectations
3) For a probability space $(\Omega, \mathcal{F}, P)$, every event
   a) is a subset of Ω
   b) is a set in the collection $\mathcal{F}$
   c) has a probability
   d) all of the above
4) In a probability triplet $(\Omega, \mathcal{F}, P)$, if $\mathcal{F}$ is the set of all subsets of Ω then
   a) all functions $X : \Omega \to \mathbb{R}$ are random variables.
   b) we cannot define random variables
   c) all events have one outcome
   d) some events are not in $\mathcal{F}$
5) If X is a random variable then $\{X \leq 5.6\}$ is
   a) an event

[3] More generally, if $X_1, \ldots, X_n$ are random variables and $h : \mathbb{R}^n \to \mathbb{R}$ is a "Borel measurable function" then $Y = h(X_1, ..., X_n)$ is a random variable. The precise definition of a *Borel measurable function* $h : \mathbb{R}^n \to \mathbb{R}$ requires the concept of the Borel sigma algebra $\mathcal{B}(\mathbb{R}^n)$ and is omitted for brevity. It can be shown that all functions $h : \mathbb{R}^n \to \mathbb{R}$ with at most finite or countably many points of discontinuity are Borel measurable.

b) an outcome
c) a real number
d) a function of ω

6) If $\{X \leq 5.6\}$ is an event, then $\{X > 5.6\}$ is an event because
a) all subsets of Ω are events
b) all events are subsets of Ω
c) the number 5.6 is special
d) the complement of an event is also an event
7) If $\{X \leq 5.6\}$ and $\{Y \leq 8.3\}$ are both events then $\{X \leq 5.6\} \cap \{Y \leq 8.3\}$ is an event because
a) the intersection of two events is also an event
b) the union of events is also an event
c) all subsets of Ω are events
d) $8.3 - 5.6 = 2.7$
8) If A is an event and $P[A] = 0$ then
a) A must be the empty set
b) A is not necessarily the empty set
c) We can apply the central limit theorem to the set A
d) We can take expectations of the set A
9) If X and Y are random variables on some probability space, then $X + Y$ is a random variable because
a) a sum of two random variables is again a random variable
b) a linear combination of two random variables is again a random variable
c) a continuous function of two random variables is again a random variable
d) all of the above
10) If X and Y are random variables that have the same CDF function, then they must be independent.
a) True
b) False
11) If Y and Z are random variables then $\{Y + e^Z \leq Z\} \in \mathcal{F}$ because
a) all subsets of Ω are events
b) $Y + e^Z - Z$ is a random variable and hence satisfies the measurability property (1)
c) Y is a function of Z
d) Y and Z are functions of ω
12) Let $\{X_i\}_{i=1}^{\infty}$ be independent and identically distributed (i.i.d.) random variables on some probability space. If the probability experiment produces a particular outcome $\omega^* \in \Omega$ then
a) the first random variable takes the value $X_1(\omega^*)$ but the values of $\{X_2, X_3, X_4, \ldots\}$ remain unknown
b) the random variables take the values $\{X_1(\omega^*), X_2(\omega^*), X_3(\omega^*), X_4(\omega^*), \ldots\}$
c) the single-outcome set $\{\omega^*\}$ must be an event
d) if $\{\omega^*\} \in \mathcal{F}$ then $P[\{\omega^*\}] > 0$

F. Summarizing exercises 2

Fix a probability space $(\Omega, \mathcal{F}, P)$. Let X, Y, Z be random variables on this space.

1) Does X necessarily have a CDF?
2) Do X and Y necessarily have a joint CDF $F_{X,Y}$?
3) Do X, Y, Z necessarily have a joint CDF $F_{X,Y,Z}$?
4) Is $4X - Y + 2.1Z$ a random variable?
5) Is $X + 2Y + 8Z$ a random variable?
6) Is $X + e^Y$ a random variable?
7) Is $\frac{YZ}{X^2+1}$ a random variable?
8) Is $\{X \leq 3\} \in \mathcal{F}$?
9) Is $\{X > 3\} \in \mathcal{F}$?
10) Is $\{X \in (2, 5]\} \in \mathcal{F}$?
11) Is $\{\frac{1}{3}(X + Y + Z) \leq 8\} \in \mathcal{F}$?
12) Is $\{e^X + \cos(Y + Z) > 2X\} \in \mathcal{F}$?
13) Is $\{\cos(X + Y) \leq 0\} \cup \{e^{4(X+Z)} > 3\} \in \mathcal{F}$?
14) Is $\cap_{i=1}^{\infty}\{\cos(iX + Y) \leq \frac{Z}{i}\} \in \mathcal{F}$?
15) Fix $A \in \mathcal{F}$ and define the *indicator function* $1_A : \Omega \to \mathbb{R}$ by

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{else} \end{cases} .$$

Is 1_A a random variable?

16) For $A \in \mathcal{F}$, is $X1_A + Y$ a random variable?
17) For $A \in \mathcal{F}$ and $B \in \mathcal{F}$, is $X1_A + e^Y 1_B + XY$ a random variable?
18) Fix $C \subseteq \Omega$ such that $C \notin \mathcal{F}$. Is the indicator function 1_C a random variable?
19) Argue that for any $a, b \in \mathbb{R}$ we have $\{X \in (a, b]\} \in \mathcal{F}$.
20) Argue that $\{X = 2.3\} \in \mathcal{F}$ by observing $\{X = 2.3\} = \cap_{i=1}^{\infty}\{X \in (2.3 - 1/i, 2.3]\}$.
21) Is $\{X < 2.3\} \in \mathcal{F}$?
22) Is $\{X \in \{2.3, 4.1, 9\}\} \in \mathcal{F}$?
23) Is $\{XY \leq 8\} \cup \{XY = 9.1\} \in \mathcal{F}$?
24) Is $\{XY \leq 7Z\} \cap \{YZ > 4\}^c \in \mathcal{F}$?
25) Is $\cup_{i=1}^{\infty}\{X = 1/i\} \in \mathcal{F}$?
26) Suppose X is a random variable that has a continuous CDF. Show that $P[X = x] = 0$ for all $x \in \mathbb{R}$. Hint: Fix $x \in \mathbb{R}$ and observe $\{X = x\} \subseteq \{X \in (x - \delta, x]\}$ for all $\delta > 0$.
27) Suppose $(\Omega, \mathcal{F}, P)$ is a probability space such that all single-outcome subsets of the sample space are events, that is, $\{\omega\} \in \mathcal{F}$ for all $\omega \in \Omega$. Suppose there is a random variable $X : \Omega \to \mathbb{R}$ that has a continuous CDF.
    a) Show that $P[\{\omega\}] = 0$ for all $\omega \in \Omega$.
    b) Show that Ω is an uncountably infinite set.
28) Fix probability space $(\Omega, \mathcal{F}, P)$. Suppose that $\{\omega\} \in \mathcal{F}$ for all $\omega \in \Omega$. Suppose $\{A_i\}_{i=1}^{\infty}$ is a sequence of mutually independent events such that $P[A_i] = 1/2$ for all $i \in \{1, 2, 3, ...\}$.
    a) Show that $P[\{\omega\}] = 0$ for all $\omega \in \Omega$.
    b) Show that Ω is an uncountably infinite set.
29) Fix probability space $(\Omega, \mathcal{F}, P)$. For each $\omega \in \Omega$ define
$$p(\omega) = \inf_{A \in \mathcal{F}: \omega \in A}\{P[A]\}$$
    a) In the special case when $\{\omega\} \in \mathcal{F}$ for all $\omega \in \Omega$, show that $P[\{\omega\}] = p(\omega)$ for all $\omega \in \Omega$.
    b) In the general case, show that for all $\omega \in \Omega$ there is an $A \in \mathcal{F}$ such that $\omega \in A$ and $P[A] = p(\omega)$.

G. Conditioning and expectation

Basic probability concepts such as a *partition of the sample space*, conditional probability, expectation, and variance are reviewed in Appendix E. This includes the *law of total probability*, the *law of total expectation*, and *linearity of expectation*:

$$P[A] = \sum_{i=1}^{\infty} P[A|B_i]P[B_i]$$
$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{E}[X|B_i] P[B_i]$$
$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i]$$

where X is a random variable with finite expectation; A is an event; $\{B_i\}_{i=1}^{\infty}$ is a collection of events that partition the sample space; $X_1, ..., X_n$ are (possibly dependent) random variables with finite expectations; $a_1, ..., a_n$ are real numbers.

H. Indicator random variable

Let A be an event. Define the binary-valued random variable 1_A by

$$1_A = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{else} \end{cases}$$

In particular, $1_A(\omega) = 1$ if and only if $\omega \in A$. The random variable 1_A is called an *indicator random variable* or *indicator function*.

I. Bernoulli trials

Fix $p \in (0, 1]$. Let $\{A_i\}_{i=1}^{\infty}$ be a sequence of mutually independent events with

$$P[A_i] = p \quad \forall i \in \{1, 2, 3, ...\}$$

This type of sequence is called *i.i.d. Bernoulli trials*. Intuitively we can imagine a sequence of independent "trials" or "mini-experiments," where each trial i can either be *successful* or *not successful*. The event A_i is the event that trial i is successful. That is, trial i is called a *success* if and only if event A_i is true.

- Bernoulli random variable ($X \sim Bern(p)$): Define X as an indicator function that is 1 if and only if the first Bernoulli trial is successful:
$$X = 1_{A_1}$$
Then
$$\begin{aligned} &X \in \{0, 1\} \\ &P[X = 1] = p, P[X = 0] = 1 - p \\ &\mathbb{E}[X] = p \\ &Var(X) = p(1 - p) \end{aligned}$$
- Geometric random variable ($X \sim Geom(p)$): Define X as the number of trials until the first success. Then[4]
$$\begin{aligned} &X \in \{1, 2, 3, ...\} \\ &P[X = k] = p(1 - p)^{k-1} \quad \forall k \in \{1, 2, 3, ...\} \\ &\mathbb{E}[X] = 1/p \\ &Var(X) = \frac{1 - p}{p^2} \end{aligned}$$
- Binomial random variable ($X \sim Binom(n, p)$): Fix n as a positive integer. Let X be the number of successes within the first n trials:
$$X = \sum_{i=1}^{n} 1_{A_i}$$
Then
$$\begin{aligned} &X \in \{0, 1, 2, ..., n\} \\ &P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad \forall k \in \{0, 1, ..., n\} \\ &\mathbb{E}[X] = np \\ &Var(X) = np(1 - p) \end{aligned}$$

J. *Computing expectations*

Assume:

- $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be measurable functions.
- X and Y are discrete random variables that take values in the finite or countably infinite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively.
- U and V are continuous random variables with joint PDF $f_{U,V}(u, v)$ and marginal PDFs $f_U(u)$ and $f_V(v)$.
- $\mathbb{E}[h(X)]$, $\mathbb{E}[g(X, Y)]$, $\mathbb{E}[h(U)]$, $\mathbb{E}[g(U, V)]$, $\mathbb{E}[g(X, V)]$ are finite.

[4]One might view X as an *extended random variable* $X : \Omega \to \mathbb{R} \cup \{\infty\}$ if there are some outcomes $\omega \in \Omega$ for which $X(\omega) = \infty$ (meaning there are *never any successes*). These outcomes, taken together, have collective probability $P[X = \infty] = 1 - \sum_{i=1}^{\infty} P[X = i] = 1 - \sum_{i=1}^{\infty} p(1 - p)^{i-1} = 0$. Alternatively, one could modify the sample space Ω by removing all outcomes $\omega \in \{X = \infty\}$. This does not change any probability or expectation because $P[X = \infty] = 0$. In general, if there is a single event A such that $P[A] = 0$, one can remove all outcomes in A from the sample space as follows: Define the new sample space $\tilde{\Omega} = \Omega \setminus A$, the new sigma algebra of events $\tilde{\mathcal{F}} = \{B \in \mathcal{F} : B \cap A = \phi\}$, the new probability measure $\tilde{P} : \tilde{\Omega} \to [0, 1]$ given by $\tilde{P}[B] = P[B]$ for all $B \in \tilde{\mathcal{F}}$. One can verify $\tilde{\mathcal{F}}$ is a sigma algebra, $\tilde{P}$ satisfies the three axioms of probability, and all remaining events have the same probability as before. Note that removing all outcomes from a single zero-probability event is not the same as removing all outcomes that are contained in *some* zero-probability event (the latter might remove *everything*).

It can be shown:

$$\begin{aligned}
\mathbb{E}\left[h(X)\right] &= \sum_{x\in\mathcal{X}} h(x)P[X=x] \\
\mathbb{E}\left[g(X,Y)\right] &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(x,y)P[X=x,Y=y] \\
\mathbb{E}\left[h(U)\right] &= \int_{-\infty}^{\infty} h(u)f_U(u)du \\
\mathbb{E}\left[g(U,V)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u,v)f_{U,V}(u,v)dudv \\
\mathbb{E}\left[g(X,V)\right] &= \sum_{x\in\mathcal{X}}\left[\int_{-\infty}^{\infty} g(x,v)f_{V|X=x}(v|x)dv\right]P[X=x] \quad (2)\\
&= \int_{-\infty}^{\infty}\left[\sum_{x\in\mathcal{X}} g(x,v)P[X=x|V=v]\right]f_V(v)dv \quad (3)
\end{aligned}$$

If $\{B_i\}_{i=1}^{\infty}$ are events that partition the sample space and W is any random variable with finite expectation then

$$\mathbb{E}\left[W\right] = \sum_{i=1}^{\infty}\mathbb{E}\left[W|B_i\right]P[B_i]$$

Similarly if W has finite expectation:

$$\begin{aligned}
\mathbb{E}\left[W\right] &= \sum_{x\in\mathcal{X}}\mathbb{E}\left[W|X=x\right]P[X=x] \\
\mathbb{E}\left[W\right] &= \int_{-\infty}^{\infty}\mathbb{E}\left[W|U=u\right]f_U(u)du \\
\mathbb{E}\left[W\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbb{E}\left[W|U=u,V=v\right]f_{U,V}(u,v)dudv \\
\mathbb{E}\left[W\right] &= \sum_{x\in\mathcal{X}}\left[\int_{-\infty}^{\infty}\mathbb{E}\left[W|X=x,V=v\right]f_{V|X=x}(v|x)dv\right]P[X=x] \quad (4)\\
&= \int_{-\infty}^{\infty}\left[\sum_{x\in\mathcal{X}}\mathbb{E}\left[W|X=x,V=v\right]P[X=x|V=v]\right]f_V(v)dv \quad (5)
\end{aligned}$$

The above formulas can be applied, for example, to compute $\mathbb{E}\left[f(R,S)\right]$ (for some given measurable function f and some given random variables R,S) by defining $W=f(R,S)$ and using the fact that a measurable function of a finite number of random variables is again a random variable.

The formulas (2), (3), (4), (5) are given in the above form for simplicity. They implicitly assume the conditional expectations are well defined and that summations and integrations are taken only over regions of positive support. For example, it is understood that the formulas (2) and (4) can be modified to $\sum_{x\in\mathcal{X}:P[X=x]>0}$. The formulas can be formally derived by using, for any event A such that $P[A]>0$ and any $v\in\mathbb{R}$ such that the PDF of V is continuous at v and $f_V(v)>0$:

$$P[A|V=v] = \frac{f_{V|A}(v)P[A]}{f_V(v)} \quad , \quad f_{V|A}(v) = \frac{P[A|V=v]f_V(v)}{P[A]}$$

and so when X is discrete and $P[X=x]>0$:

$$f_{V|X=x}(v|x) = \frac{P[X=x|V=v]f_V(v)}{P[X=x]} \quad (6)$$

For example, temporarily ignoring technical issues of "divide by zero" and "PDF continuity," we have

$$\begin{aligned}
\mathbb{E}\left[g(X,V)\right] &= \sum_{x\in\mathcal{X}} \mathbb{E}\left[g(X,V)|X=x\right] P[X=x] \\
&= \sum_{x\in\mathcal{X}} \mathbb{E}\left[g(x,V)|X=x\right] P[X=x] \\
&= \sum_{x\in\mathcal{X}} \left[\int_{-\infty}^{\infty} g(x,v) f_{V|X=x}(v|x)dv\right] P[X=x] \quad (7) \\
&\overset{(a)}{=} \sum_{x\in\mathcal{X}} \left[\int_{-\infty}^{\infty} g(x,v)\frac{P[X=x|V=v]f_V(v)}{P[X=x]}dv\right] P[X=x] \\
&= \sum_{x\in\mathcal{X}} \left[\int_{-\infty}^{\infty} g(x,v)P[X=x|V=v]f_V(v)dv\right] \\
&\overset{(b)}{=} \int_{-\infty}^{\infty}\left[\sum_{x\in\mathcal{X}} g(x,v)P[X=x|V=v]\right] f_V(v)dv \quad (8)
\end{aligned}$$

where equality (a) uses (6); equality (b) switches the order of summation and integration and is formally justified by the Fubini-Tonelli theorem of measure theory (and the fact that $\mathbb{E}\left[g(X,V)\right]$ is finite). Observe that (7) is identical to formula (2); (8) is identical to formula (3). The derivation of formulas (4) and (5) is similar. It should be emphasized that the above derivation is "stylized" because it ignores conditions for (6) to make sense ($P[X=x]>0$, $f_V(v)>0$, the PDF f_V is continuous at v). In most cases when these formulas are applied, the PDFs or conditional PDFs are continuous or piecewise continuous over a region of positive support. The integrals and summations can be taken over appropriate terms and intervals.

The formula (6) treats the case when X is discrete type and V is continuous type. When U,V are continuous with joint PDF $f_{U,V}(u,v)$ the formula becomes $f_{V|U}(v|u)=\frac{f_{U|V}(u|v)f_V(v)}{f_U(u)}$ via the definition $f_{U|V}(u|v)=\frac{f_{U,V}(u,v)}{f_V(v)}$.

K. Exercises (computing expectations)

The following multiple choice problems can be solved by computer. For example, to verify that $\sum_{i=1}^{\infty} i(1/2)^i = 2$ see:
`https://www.wolframalpha.com/input?i=sum+i*%281%2F2%29%5Ei%2C+i%3D1..%5Cinfty`
The problems ask you to compute expectations. Assume X,Y,U,V are mutually independent random variables and

- X and Y are discrete type with $X\sim Geom(1/2)$, $Y\sim Unif\{1,2,3\}$.
- U and V are continuous type with $U\sim Exp(2)$ and $V\sim Unif[0,1]$.

1) $\mathbb{E}\left[X\right]$
   a) 3
   b) 2
   c) 2.01
   d) -4

2) $\mathbb{E}\left[XY\right]$
   a) 4
   b) 3.01
   c) 3
   d) 0.5

3) $\mathbb{E}\left[X\cos(X)\right]$
   a) 0.8763390100
   b) -0.3222562303
   c) 2.5489534553
   d) 0.7427489902

4) $\mathbb{E}\left[\cos(X+Y)\right]$
   a) 0.5001128203
   b) -0.0034884486
   c) 0.4784639274
   d) -0.3820557339

5) $\mathbb{E}\left[\frac{X+\cos(XY^2)}{1+X^2}\right]$
   a) 0.2117893563
   b) -1.3378566408
   c) 0.3256249016
   d) 0.3899440031

6) $\mathbb{E}\left[U\right]$
   a) 3
   b) 2
   c) 1/2
   d) $1/\pi$

7) $\mathbb{E}\left[U^2 + U\right]$
   a) 3
   b) 4
   c) 5
   d) 1

8) $\mathbb{E}\left[U\cos(U)\sin(U^2)\right]$
   a) 0.334671
   b) 0.224671
   c) 0.114671
   d) 0.004671

9) $\mathbb{E}\left[\cos(U+V)\right]$
   a) 0.48930
   b) 0.58930
   c) 0.68930
   d) 0.10334

10) $\mathbb{E}\left[\frac{U}{1+V^2}\right]$
   a) 0.39270
   b) 0.12124
   c) 0.00001
   d) 0.29938

11) $\mathbb{E}\left[\cos(X+V)\right]$
   a) -0.11874
   b) -0.24863
   c) 0.34998
   d) 0

12) $\mathbb{E}\left[\frac{X}{1+V^2}\right]$
   a) 1.5708
   b) 1.9903
   c) 1.1001
   d) 2.3202

13) Suppose $\mathbb{E}\left[W|X=k\right] = k^2$ for all relevant k. $\mathbb{E}\left[W\right] = ?$
   a) 5.5
   b) 6
   c) 6.5
   d) 7

14) Suppose $\mathbb{E}\left[W|X=k\right] = \cos(k+4.5)$ for all relevant k. $\mathbb{E}\left[W\right] = ?$
   a) 0.101136
   b) 0.205982

c) -0.490835
d) 0.573531

15) Suppose $\mathbb{E}\left[W^2|X=n,Y=m\right]=\exp(-(n+m))$ for all relevant n,m. $\mathbb{E}\left[W^2\right]=$?
a) 0.2098740
b) 0.0415488
c) 0.0539852
d) 0.0659552

16) Suppose $\mathbb{E}\left[W|U=u,V=v\right]=uv$ for all relevant u,v. $\mathbb{E}\left[W\right]=$?
a) 1/2
b) 1/4
c) 1/8
d) 1/16

17) Suppose $\mathbb{E}\left[W|U=u,V=v\right]=\cos(u+v^2)$ for all relevant u,v. $\mathbb{E}\left[W\right]=$?
a) 0.599512
b) 0.699512
c) 0.399512
d) 0.299512

18) Suppose $\mathbb{E}\left[e^W|X=n,V=v\right]=nv$ for all relevant n,v. $\mathbb{E}\left[e^W\right]=$?
a) 0.89887
b) 0.78994
c) 1
d) 0.93032

19) Suppose $\mathbb{E}\left[W|X=n,V=v\right]=\cos(n+v)$ for all relevant n,v. $\mathbb{E}\left[W\right]=$?
a) 0.36561
b) -0.18879
c) 0.19454
d) -0.24863

20) Give a formula for $\mathbb{E}\left[r(U,V,W)\right]$ given a measurable function r and a joint PDF $f_{U,V,W}(u,v,w)$.

21) Let A be an event and suppose $P[A|X=x]$ is known for all relevant x. Give a formula for $P[A]$. Design and solve a problem of this type.

22) Let A be an event and suppose $P[A|X=x,Y=y]$ is known for all relevant x,y. Give a formula for $P[A]$. Design and solve a problem of this type.

23) Let A be an event and suppose $P[A|U=u]$ is known for all relevant u. Give a formula for $P[A]$. Design and solve a problem of this type.

24) Let A be an event and suppose $P[A|U=u,V=v]$ is known for all relevant u,v. Give a formula for $P[A]$. Design and solve a problem of this type.

25) Let A be an event and suppose $P[A|X=x,V=v]$ is known for all relevant x,v. Give a formula for $P[A]$. Design and solve a problem of this type.

L. Make your own examples

1) Give an example of random variables X,Y that are dependent.
2) Give an example of random variables X,Y such that $X\sim Unif[0,1]$, Y is continuous type, and $\mathbb{E}\left[Y|X=x\right]=3x$ for all $x\in[0,1]$. Are X and Y independent?
3) Give an example of random variables X,Y such that $X\sim Unif[0,1]$, Y is discrete type, and $\mathbb{E}\left[Y|X=x\right]=3x$ for all $x\in[0,1]$. Are X and Y independent?
4) Give an example of random variables X,Y such that $X\sim Unif[0,1]$, and $\mathbb{E}\left[Y|X=x\right]=x^2$ for all $x\in[0,1]$. Are X and Y independent?

5) Give an example of a random variable X such that $\mathbb{E}[X]$ is finite but $\mathbb{E}\left[X^2\right] = \infty$.
6) Give an example of a random variable such that $\mathbb{E}[X] = DNE$ (Does Not Exist).
7) Give an example of a continuous type random variable with $\mathbb{E}[X] = 1$ and $Var(X) = 3$.
8) Give an example of a discrete type random variable with $\mathbb{E}[X] = 1$ and $Var(X) = 3$.
9) Give an example of a random variable X with $\mathbb{E}[X] = 0$, $\mathbb{E}\left[X^2\right] = 8$, $\mathbb{E}\left[X^4\right] = 65$.
10) Is it possible to construct a random variable X with $\mathbb{E}[X] = 0$, $\mathbb{E}\left[X^2\right] = 8$, $\mathbb{E}\left[X^4\right] = 60$?
11) Give an example of *dependent* random variables X, Y such that $X \in \{0, 1\}$ and $\mathbb{E}[Y|X = x] = 0$ for all $x \in \{0, 1\}$.
12) Give an example of random variables X, Y that are uncorrelated but not independent.
13) Give an example of something interesting that involves random variables.

III. Convergence

A. Convergence surely, almost surely, and in probability

Fix a probability space $(\Omega, \mathcal{F}, P)$. Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of random variables. Fix $c \in \mathbb{R}$.

Definition 1: We say $Y_n \to c$ *surely* if

$$\lim_{n\to\infty} Y_n(\omega) = c \quad \forall \omega \in \Omega$$

Definition 2: We say $Y_n \to c$ *almost surely* (also called *convergence with probability 1*) if

$$P\left[\left\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = c\right\}\right] = 1$$

Definition 3: We say $Y_n \to c$ *in probability* if for all $\epsilon > 0$ we have

$$\lim_{n\to\infty} P[|Y_n - c| \geq \epsilon] = 0$$

The set $\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = c\}$ used in Definition 2 is formally shown to be in $\mathcal{F}$ in Appendix D. With respect to these types of convergence, in Appendix C it is shown that:

$$\text{surely} \implies \text{almost surely} \implies \text{in probability}$$

Convergence to a constant c can be used to define convergence to a random variable Y. We say $Y_n \to Y$ surely if and only if $(Y_n - Y) \to 0$ surely. Similarly, $Y_n \to Y$ almost surely if and only if $(Y_n - Y) \to 0$ almost surely; $Y_n \to Y$ in probability if and only if $(Y_n - Y) \to 0$ in probability. It can be shown that if Y_n converges to Y in probability then a property called *convergence in distribution* holds: $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$ for all $y \in \mathbb{R}$ at which F_Y is continuous, where $F_{Y_n}(y) = P[Y_n \leq y]$ and $F_Y(y) = P[Y \leq y]$ are the CDF functions for Y_n and Y.

B. Law of large numbers (LLN)

Random variables $\{X_i\}_{i=1}^{\infty}$ are said to be *independent and identically distributed* (i.i.d.) if they are mutually independent and if all random variables have the same marginal CDF, so that

$$F_{X_i}(x) = F_{X_1}(x) \quad \forall x \in \mathbb{R}, \forall i \in \{1, 2, 3, \ldots\}$$

Theorem 1: (Law of Large Numbers – LLN) If $\{X_i\}_{i=1}^{\infty}$ are i.i.d. random variables with finite mean μ (so $\mathbb{E}[X_i] = \mu$ for all $i \in \{1, 2, 3, \ldots\}$), then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu \quad \text{almost surely}$$

Proof: A proof for the special case when the random variables have a finite variance is given in Appendix C. A full proof that allows $Var(X_i) = \infty$ is given in [17]. □

The above theorem requires the random variables X_i to have a finite mean, but allows them to have a possibly infinite variance. This version of the LLN is often called the *strong law of large numbers* because it treats convergence almost surely. The almost sure convergence in Theorem 1 means that if we define A as the set of all outcomes in the sample space for which the limit converges to μ, then $P[A] = 1$. That is

$$A = \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = \mu\right\}$$

and $P[A] = 1$. Defining $A^c = \Omega \setminus A$ as the complement of set A, it follows that $P[A^c] = 0$. The set A^c contains all outcomes for which the limit either does not exist, or the limit exists but is not equal to μ. The set A^c may contain uncountably many outcomes, so it is remarkable that the LLN ensures $P[A^c] = 0$.

Appendix C shows that convergence almost surely implies the weaker *convergence in probability*: For all $\epsilon > 0$ we have

$$\lim_{n\to\infty} P\left[\left|\mu - \frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \epsilon\right] = 0 \tag{9}$$

In many applications, the i.i.d. random variables $\{X_i\}_{i=1}^{\infty}$ not only have finite mean μ, but they have finite variance σ^2 (so $\mathbb{E}[X_i] = \mu$ and $Var(X_i) = \sigma^2$ for all $i \in \{1, 2, 3, \ldots\}$). In this case it holds by *linearity of expectation* and basic properties of variance that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mu \quad \forall n \in \{1, 2, 3, \ldots\}$$

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) \overset{(a)}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n} \quad \forall n \in \{1, 2, 3, \ldots\}$$

where step (a) uses the fact that the variance of a sum of pairwise uncorrelated random variables is the sum of variances. It follows that the mean of $\frac{1}{n}\sum_{i=1}^{n} X_i$ is always μ, while its variance converges to 0 as $n \to \infty$. This provides a great deal of intuition for the LLN. Further, the fact that the variance converges to zero can be used, together with the Markov/Chebyshev inequality, to directly prove the weak convergence result given in (9). More work is required to prove the stronger result for almost sure convergence given in Theorem 1 (see Appendix C and [17]).

C. Applying the LLN twice (motivation for renewal theory)

If A and B are events that satisfy $P[A] = 1$ and $P[B] = 1$, the axioms of probability can be used to show that $A \cap B$ is an event that satisfies $P[A \cap B] = 1$. This can be immediately used in the following important example: Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables with finite mean 5.7. Let $\{T_i\}_{i=1}^{\infty}$ be i.i.d. random variables with mean 17.1. So

$$\mathbb{E}[X_i] = 5.7 \quad \forall i \in \{1, 2, 3, \ldots\}$$

$$\mathbb{E}[T_i] = 17.1 \quad \forall i \in \{1, 2, 3, \ldots\}$$

Further assume each T_i is *surely positive*, so that $T_i(\omega) > 0$ for all $i \in \{1, 2, 3, \ldots\}$ and all $\omega \in \Omega$. In this example, the random variables $\{X_i\}_{i=1}^{\infty}$ are assumed to be i.i.d. amongst themselves; the random variables $\{T_i\}_{i=1}^{\infty}$ are i.i.d. amongst themselves. However, there can be arbitrary dependencies between the X_i and T_j variables. For example, we might have $T_i = \cos(X_i) + 5$ for all $i \in \{1, 2, 3, \ldots\}$.

For each positive integer n, define the random variable Z_n by

$$Z_n = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} T_i}$$

The random variables T_i are assumed to be positive to avoid the undefined case of division by zero.

Does Z_n converge to anything? If so, does it converge almost surely? To answer this question, let us first fix a particular $\omega^* \in \Omega$ to see what happens. For this fixed ω^* we know that $\{X_i(\omega^*)\}_{i=1}^{\infty}$ and $\{T_i(\omega^*)\}_{i=1}^{\infty}$ and $\{Z_n(\omega^*)\}_{n=1}^{\infty}$ are just deterministic sequences of real numbers. Now suppose our fixed ω^* *just so happens* to be an outcome that satisfies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i(\omega^*) = 5.7 \tag{10}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} T_i(\omega^*) = 17.1 \tag{11}$$

Then we can conclude:

$$\begin{aligned}\lim_{n\to\infty} Z_n(\omega^*) &\overset{(a)}{=} \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i(\omega^*)}{\sum_{i=1}^{n} T_i(\omega^*)} \\ &\overset{(b)}{=} \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^{n} X_i(\omega^*)}{\frac{1}{n}\sum_{i=1}^{n} T_i(\omega^*)} \\ &\overset{(c)}{=} \frac{5.7}{17.1}\end{aligned} \tag{12}$$

where in (12), the equality (a) uses the definition of Z_n; the equality (b) multiplies and divides by $1/n$; the equality (c) uses the basic calculus fact that the ratio of well defined limits is equal to the limit of the ratio. Therefore, if our particular ω^* just so happens to satisfy *both* (10) *and* (11), then we can conclude

$$\lim_{n\to\infty} Z_n(\omega^*) = \frac{5.7}{17.1}$$

Of course, in a real problem, the outcome $\omega \in \Omega$ is chosen *randomly* by nature (according to the probability measure P). How likely is it that the particular ω chosen by nature satisfies *both* (10) and (11)? The LLN applied to the i.i.d. sequence $\{X_i\}_{i=1}^{\infty}$ ensures that our randomly selected ω satisfies (10) *with probability 1*. The LLN applied to the i.i.d. sequence $\{T_i\}_{i=1}^{\infty}$ ensures that our randomly selected ω satisfies (11) *with probability 1*. Hence, our randomly selected ω satisfies *both* (10) and (11) with probability 1. We formalize this argument in the following claim.

Claim: We have

$$Z_n \to \frac{5.7}{17.1} \quad \text{almost surely}$$

Proof: Define the following events A, B, C:[5]

$$A = \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = 5.7\right\}$$
$$B = \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} T_i(\omega) = 17.1\right\}$$
$$C = \left\{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = \frac{5.7}{17.1}\right\}$$

The LLN ensures $P[A] = 1$ and $P[B] = 1$. Thus, $P[A \cap B] = 1$. However, if $\omega \in A \cap B$, we know (by the fact that the limit of a ratio is the ratio of limits) that $\omega \in C$. Indeed, just repeat the argument (12) but change ω^* to ω. It follows that

$$A \cap B \subseteq C$$

Thus

$$\underbrace{P[A \cap B]}_{1} \leq P[C]$$

Since we already know the left-hand-side of the above inequality is 1, and $P[C]$ cannot be larger than 1, it follows that $P[C] = 1$. $\square$

We have made good use of the following probability fact:

$$(P[A] = 1 \text{ and } P[B] = 1) \implies (P[A \cap B] = 1)$$

The following is a generalized version of this fact that may be useful in related contexts:

Theorem 2: Let $\{A_i\}_{i=1}^{\infty}$ be a countably infinite sequence of events.

a) If $P[A_i] = 0$ for all $i \in \{1, 2, 3, \ldots\}$ then $P[\cup_{i=1}^{\infty} A_i] = 0$.

b) If $P[A_i] = 1$ for all $i \in \{1, 2, 3, \ldots\}$ then $P[\cap_{i=1}^{\infty} A_i] = 1$.

Proof: To prove part (a), suppose $P[A_i] = 0$ for all $i \in \{1, 2, 3, \ldots\}$. The union bound implies

$$P[\cup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} P[A_i] = \sum_{i=1}^{\infty} 0 = 0$$

This proves part (a).

To prove part (b), suppose $P[A_i] = 1$ for all $i \in \{1, 2, 3, \ldots\}$. It suffices to prove that

$$P[(\cap_{i=1}^{\infty} A_i)^c] = 0$$

We have by DeMorgan's law

$$(\cap_{i=1}^{\infty} A_i)^c = \cup_{i=1}^{\infty} A_i^c$$

Thus

$$P[(\cap_{i=1}^{\infty} A_i)^c] = P[\cup_{i=1}^{\infty} A_i^c] \leq \sum_{i=1}^{\infty} P[A_i^c] = 0$$

which uses the fact that $P[A_i^c] = 0$ for all $i \in \{1, 2, 3, \ldots\}$. $\square$

[5] A representation in (64) of Appendix D formally proves the sets A, B, C are events.

D. *Exercises*

For the following exercises, assume

- $\{X_i\}_{i=1}^{\infty}$ are i.i.d. $Bern(1/2)$
- $\{Y_i\}_{i=1}^{\infty}$ are i.i.d. $N(0,1)$
- $\{Z_i\}_{i=1}^{\infty}$ are i.i.d. $Geom(1/5)$
- $\{W_i\}_{i=1}^{\infty}$ are i.i.d. $Bern(p)$
- The sequences $\{X_i\}, \{Y_i\}, \{Z_i\}, \{W_i\}$ may depend on each other

1) The expectation of an indicator function is
   a) 0
   b) the variance of its inverse
   c) an outcome
   d) the probability the indicator function is 1

2) For insight into how closely $\frac{1}{100}\sum_{i=1}^{100} W_i$ approximates p we can
   a) discard half the samples
   b) compute $Var(\frac{1}{100}\sum_{i=1}^{100} W_i)$
   c) assume $W_i \sim N(0,1)$.
   d) use the Slytherin theorem

3) The probability that both $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = 1/2$ and $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = 0$ is
   a) Unknown
   b) 1
   c) 0
   d) $1/4$

4) With prob 1 $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \cos(X_i) =$
   a) 0
   b) 0.2309
   c) -0.3480
   d) 0.7702

5) With prob 1 $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (Y_i + Z_i) =$
   a) 5
   b) 25
   c) 1/10
   d) 1

6) With prob 1 $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Z_i} =$
   a) 1/20
   b) 4
   c) 1/10
   d) Unknown

7) With prob 1 $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} (X_i + Y_i)}{\sum_{i=1}^{n} (1+Y_i^2)} =$
   a) 2
   b) 1/4
   c) 1/8
   d) 1/5

8) With prob 1 $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i =$
   a) 1/2
   b) 1/3
   c) 2
   d) Unknown

9) With prob 1 $\lim_{n\to\infty} \frac{1}{8n}\sum_{i=1}^{n} X_{2i} =$

a) 1/8
b) 1/32
c) 1/2
d) 1/16

10) With prob 1 $\lim_{n\to\infty} \frac{\sum_{i=1}^n e^{X_i}}{\sum_{i=1}^n (1+Z_i)} =$
a) 0.2099
b) 0.3099
c) 0.4099
d) 0.5099

11) $\mathbb{E}\left[\frac{1}{100}\sum_{i=1}^n Y_i\right] =$
a) 5
b) 0
c) 10
d) 2.5

12) $Var(\frac{1}{100}\sum_{i=1}^n X_i) =$
a) 1/400
b) 1/1600
c) 1/100
d) 2/350

13) $Var(\frac{1}{1000}\sum_{i=1}^{1000} Z_i) =$
a) 0.02
b) 0.15
c) 0.90
d) 1.80

14) Explain why $Var(\frac{1}{100}\sum_{i=1}^{100} X_i Z_i)$ is unknown.

E. Design your own examples

1) Fix $p \in (0,1)$. Give an example of a sequence $\{V_i\}_{i=1}^{\infty}$ of $Bern(p)$ random variables such that $\frac{1}{n}\sum_{i=1}^n V_i$ does *not* converge to p almost surely as $n \to \infty$. Explain why the example requires $p \in (0,1)$ instead of $p \in [0,1]$.

2) Fix $p \in [0,1]$. Give an example of a sequence $\{V_i\}_{i=1}^{1000}$ of $Bern(p)$ random variables such that

$$Var\left(\frac{1}{1000}\sum_{i=1}^{1000} V_i\right) = \frac{p(1-p)}{2}$$

F. Exercises related to simulation

Suppose that a randomly formed network with a certain structure is connected with some unknown probability p. To estimate p, we write a computer program that loops over iterations $i \in \{1,\dots,n\}$. At iteration i, a new network i with the desired structure is randomly generated using multiple calls to a random number generator rand(). Then, a subroutine is run to determine whether or not that particular network i is connected. Let X_i be a Bernoulli random variable that is 1 if network i is connected, and 0 else. Define $\hat{P}_n = \frac{1}{n}\sum_{i=1}^n X_i$ as an estimate of p based on these n samples.

1) $\{X_i\}_{i=1}^n$ are identically distributed because
a) They are independent
b) They are part of an infinite sequence
c) The same software is used to generate each network i
d) All of the networks $i \in \{1,...,n\}$ are connected

2) $\{X_i\}_{i=1}^n$ are mutually independent because
a) They are part of an infinite sequence
b) The same software is used to generate each network i
c) They have the same distribution

d) Each call to rand() generates a random variable that is independent of the past

3) Which theorem relates to using $\hat{P}_n$ as an estimate of p?
   a) The law of large numbers
   b) The law of total probability
   c) The law of total expectation
   d) DeMorgan's law

4) Compute $\mathbb{E}\left[\hat{P}_n\right]$ and $Var(\hat{P}_n)$ for all $n \in \{1, 2, 3, ...\}$.

5) Fix $n = 800$. Give a numerical upper bound on the mean square error of the estimate $\hat{P}_{800}$. That is, compute an upper bound on $\mathbb{E}\left[(\hat{P}_{800} - p)^2\right]$. Your bound should be a number (not a function of p).

6) Fix $n = 800$. Use the Chebyshev inequality (see Appendix F) to find a 95% confidence interval for p based on 800 samples. That is, find a numerical value $\epsilon > 0$ (that does not depend on p) so that $P[p \in [\hat{P}_{800} - \epsilon, \hat{P}_{800} + \epsilon]] \geq 0.95$.

7) Fix $n = 800$. Use the Hoeffding inequality (see Appendix F) to find a 95% confidence interval for p based on 800 samples. That is, find a numerical value $\epsilon > 0$ (that does not depend on p) so that $P[p \in [\hat{P}_{800} - \epsilon, \hat{P}_{800} + \epsilon]] \geq 0.95$. How does the size of the interval compare to the interval obtained by the Chebyshev inequality?

8) Using the Hoeffding inequality (see Appendix F), how many samples n are needed to obtain a 95% confidence interval $[\hat{P}_n - 0.01, \hat{P}_n + 0.01]$ for p?

G. *Additional exercises*

1) Fix $(\Omega, \mathcal{F}, P)$. Let $X : \Omega \to \mathbb{R}$ be a random variable. Suppose $X \sim Geom(p)$ for some $p \in [0, 1]$. Define $Y_n = 2^n 1_{\{X>n\}}$ for $n \in \{1, 2, 3, ...\}$. Show that $Y_n \to 0$ surely. Compute $\mathbb{E}[Y_n]$ and $\mathbb{E}\left[Y_n^2\right]$ for all n.

2) Fix $(\Omega, \mathcal{F}, P)$. For simplicity, assume $\{\omega\} \in \mathcal{F}$ for all $\omega \in \Omega$. Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of random variables. Show that in the special case when Ω is a finite or countably infinite set, then $X_i \to 0$ in probability if and only if $X_i \to 0$ almost surely.

IV. RENEWAL THEORY

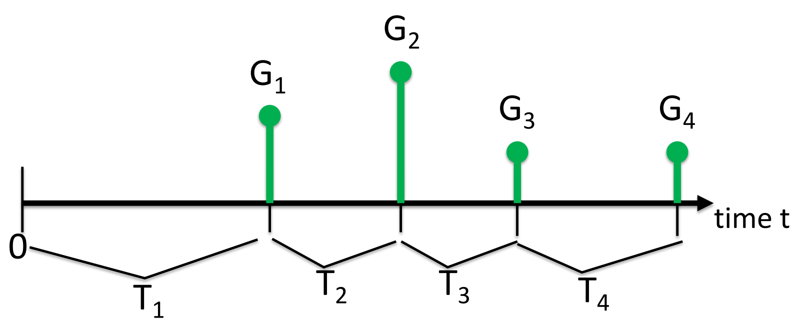

Fig. 1. A timeline showing back-to-back renewal frames with i.i.d. sizes $\{T_i\}_{i=1}^{\infty}$ and i.i.d. rewards $\{G_i\}_{i=1}^{\infty}$.

A general theory of *renewal systems* can be understood as a generalization of the example of Section III-C. Consider a stochastic system where objects arrive over a continuous timeline $t \geq 0$ (see Fig. 1). For simplicity, it is assumed that two objects cannot arrive at the same time. One can imagine that each new arrival i brings a random *reward* G_i. Fig. 1 illustrates the *inter-arrival times* $\{T_i\}_{i=1}^{\infty}$ and the *rewards* $\{G_i\}_{i=1}^{\infty}$. The inter-arrival times are assumed to surely be positive, that is

$$T_i(\omega) > 0 \quad \forall i \in \{1, 2, 3, \ldots\}, \forall \omega \in \Omega$$

In many scenarios of interest the rewards G_i are nonnegative, but the general theory allows for general rewards that can possibly take negative values. That is, the random variables G_i are allowed to take any sign (positive, negative, or zero).

An example scenario is when data packets arrive to a network at distinct instants of time. Defining G_i as the bit length of packet i means that the sum reward up to a particular time is equal to the total number of bits that have arrived to the network. Defining $G_i = 1$ for all i means that the sum reward up to a particular time is just the total number of packet arrivals up to that time (counting each packet separately, regardless of its size).

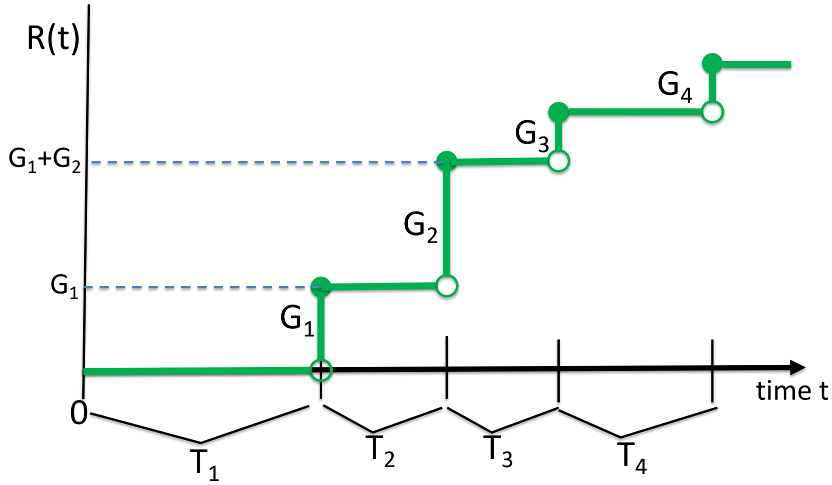

Fig. 2. A timeline showing the reward process $R(t)$ corresponding to Fig. 1. The height of each jump i is equal to the reward G_i. The corresponding $N(t)$ process is not shown, but is easily formed by making the height of each jump equal to 1 (rather than G_i).

We can define notation for the time of the ith arrival: Let Z_n be the time of arrival $n \in \{1, 2, 3, \dots\}$, so

$$\begin{aligned} Z_1 &= T_1 \\ Z_2 &= T_1 + T_2 \\ Z_3 &= T_1 + T_2 + T_3 \end{aligned}$$

and so on, so that

$$Z_n = \textstyle\sum_{i=1}^n T_i$$

For real-valued time $t \geq 0$, define $N(t)$ as the random process that counts the total number of arrivals up to and including time t. The process $N(t)$ is called a *counting process*. Unless otherwise stated, it is assumed that $N(0) = 0$ (so we imagine there are no arrivals at or before the starting time $t = 0$). The $N(t)$ function is a nondecreasing "staircase" function that increases by exactly one at the times Z_1, Z_2, Z_3 and so on. If we happen to evaluate the $N(t)$ function at the arrival times Z_n we obtain

$$N(Z_n) = n \quad \forall n \in \{1, 2, 3, \dots\}$$

which means that the number of arrivals at the time of the first arrival is 1; the number of arrivals at the time of the second arrival is 2; and so on. What if we evaluate the $N(t)$ function at a time t^* that is in between two arrivals? For example, suppose time t^* satisfies $Z_5 < t^* < Z_6$. Then exactly 5 arrivals have occured up to and including time t, so

$$N(t^*) = N(Z_5) = 5$$

The function $N(t)$ will not increase to 6 until time $t = Z_6$, at which point it will never go lower than 6.

Let $R(t)$ denote the total reward accumulated up to and including time t (see Fig. 2):

$$R(t) = \sum_{i=1}^{N(t)} G_i \quad \forall t \geq 0$$

Then $R(0) = 0$ and the $R(t)$ function is piecewise constant, increasing (or decreasing) at the time of each new arrival i according to the value of the reward G_i. Recall that there is an underlying probability space $(\Omega, \mathcal{F}, P)$ and that a particular outcome $\omega \in \Omega$ completely determines all $\{T_i(\omega)\}_{i=1}^{\infty}$ and $\{G_i(\omega)\}_{i=1}^{\infty}$ values. Therefore, a given outcome $\omega \in \Omega$ completely determines $N(t)$ and $R(t)$ for all $t \geq 0$. We can explicitly represent this dependence on ω by using extended notation $N(t, \omega)$ and $R(t, \omega)$. However, we shall hide the dependence on ω for simplicity of notation.

Definition 4: We say the process has a *time average arrival rate* λ and a *time average reward rate* α if λ and α are finite constants that satisfy

$$\lim_{t\to\infty} \frac{N(t)}{t} = \lambda \quad \text{almost surely} \tag{13}$$

$$\lim_{t\to\infty} \frac{R(t)}{t} = \alpha \quad \text{almost surely} \tag{14}$$

where we recall that the phrase "almost surely" is synonymous with "with probability 1."

Thus far we have not imposed any probability assumptions on the sequences $\{T_i\}_{i=1}^{\infty}$ or $\{G_i\}_{i=1}^{\infty}$ (such as i.i.d. assumptions). In general, the limits in (13) and (14) are not guaranteed to exist with probability 1. However, if $\{T_i\}_{i=1}^{\infty}$ are i.i.d. with a finite mean, and if $\{G_i\}_{i=1}^{\infty}$ are i.i.d. with a finite mean, then finite values of λ and α exist and can be easily calculated. For notational purposes, it shall be useful to define $T = T_1$ and $G = G_1$ so that the finite means can be expressed as $\mathbb{E}[T]$ and $\mathbb{E}[G]$.

Theorem 3: (Renewal-reward) Supppose $\{T_i\}_{i=1}^{\infty}$ are i.i.d. positive random variables with finite mean $\mathbb{E}[T] > 0$. Suppose $\{G_i\}_{i=1}^{\infty}$ are i.i.d. random variables with finite mean $\mathbb{E}[G]$. Then

a) We surely have an infinite number of arrivals:

$$\lim_{t\to\infty} N(t) = \infty \quad \text{surely}$$

b) The time average arrival rate satisfies

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[T]} \quad \text{almost surely}$$

so that $\lambda = 1/\mathbb{E}[T]$.

c) The time average rate of rewards satsifies

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[G]}{\mathbb{E}[T]} \quad \text{almost surely}$$

so that $\alpha = \mathbb{E}[G]/\mathbb{E}[T]$.

Proof: To prove part (a), since $T_i : \Omega \to (0,\infty)$ are surely positive real numbers for each i, the nth arrival comes at the finite time $\sum_{i=1}^{n} T_i$. So for any positive integer n, we have $N(t) \geq n$ whenever $t \geq \sum_{i=1}^{n} T_i$. This proves part (a).

To prove part (b), fix $t \geq T_1$. Then $N(t) \geq 1$ and t is in between two renewal times, so

$$\sum_{i=1}^{N(t)} T_i \leq t < \sum_{i=1}^{N(t)+1} T_i$$

Dividing both sides by the nonzero value $N(t)$ gives

$$\frac{1}{N(t)}\sum_{i=1}^{N(t)} T_i \leq \frac{t}{N(t)} < \left(\frac{N(t)+1}{N(t)}\right)\left(\frac{1}{N(t)+1}\sum_{i=1}^{N(t)+1} T_i\right) \tag{15}$$

The above holds for all $t \geq T_1$. Part (a) ensures $N(t) \to \infty$ surely, and so

$$\lim_{t\to\infty} \frac{N(t)+1}{N(t)} = 1 \quad \text{surely} \tag{16}$$

$$\lim_{t\to\infty} \frac{1}{N(t)}\sum_{i=1}^{N(t)} T_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} T_i = \mathbb{E}[T] \quad \text{almost surely} \tag{17}$$

$$\lim_{t\to\infty} \frac{1}{N(t)+1}\sum_{i=1}^{N(t)+1} T_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} T_i = \mathbb{E}[T] \quad \text{almost surely} \tag{18}$$

where (17) and (18) hold by the LLN. The probability that the limits in (16), (17), (18) all hold simultaneously is 1. Taking limits in (15) as $t \to \infty$ and substituting (16),(17),(18) yields

$$\mathbb{E}[T] \leq \lim_{t\to\infty} \frac{t}{N(t)} \leq \mathbb{E}[T] \quad \text{almost surely}$$

This proves $t/N(t) \to \mathbb{E}[T]$ almost surely. Since $\mathbb{E}[T] > 0$, it follows that $N(t)/t \to 1/\mathbb{E}[T]$ almost surely.

To prove (c), fix $t \geq Z_1$. Then $t > 0$, $N(t) \geq 1$, and

$$\frac{R(t)}{t} = \frac{\sum_{i=1}^{N(t)} G_i}{t} = \left(\frac{N(t)}{t}\right)\left(\frac{1}{N(t)}\sum_{i=1}^{N(t)} G_i\right) \tag{19}$$

On the other hand, part (b) implies

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[T]} \quad \text{almost surely} \tag{20}$$

while the LLN together with the fact that $N(t) \to \infty$ implies

$$\lim_{t\to\infty} \frac{1}{N(t)} \sum_{i=1}^{N(t)} G_i = \mathbb{E}[G] \quad \text{almost surely} \tag{21}$$

Recall that if $P[A] = 1$ and $P[B] = 1$ then $P[A \cap B] = 1$. Thus, the probability that the limits in (20) and (21) hold simultaneously is 1. Taking a limit in (19) as $t \to \infty$ and substituting (20) and (21) yields

$$\lim_{t\to\infty} \frac{R(t)}{t} = \left(\frac{1}{\mathbb{E}[T]}\right)\mathbb{E}[G] \qquad \text{almost surely}$$

which proves part (c). □

Theorem 3 has applications for analysis of *renewal systems*, being systems that "refresh" or "renew" or "regenerate" themselves at particular instants of time. Such systems are also called *regenerative systems*. For these systems, the times $\{T_i\}_{i=1}^{\infty}$ are called the *inter-renewal times*, so $Z_1 = T_1$ is the time of the first renewal, $Z_2 = T_1 + T_2$ is the time of the second renewal, and so on. The period of time between renewal $i-1$ and renewal i is called *renewal frame* i and has duration T_i. The value G_i is the total reward earned on frame i. At the start of renewal frame i, the system resets and repeats its behavior in an independent but identically distributed way. One can imagine computer software independently outputting a new random vector (T_i, G_i) for each frame $i \in \{1, 2, 3, \ldots\}$ by running the same subroutine but having all sources of randomness in the subroutine generated independently. Define $T = T_1$ and $G = G_1$. Each random vector (T_i, G_i) has the same distribution as (T, G), so the joint CDF satisfies

$$P[T_i \leq \tau, G_i \leq g] = F_{T,G}(\tau, g) \quad \forall (\tau, g) \in \mathbb{R}^2, \quad \forall i \in \{1, 2, 3, \ldots\}$$

Independence means that for all $i \geq 2$, knowledge of what happened in previous frames does not affect the result of frame i:

$$P[T_i \leq \tau, G_i \leq g | (G_1, T_1), \ldots, (G_{i-1}, T_{i-1})] = F_{T,G}(\tau, g) \quad \forall (\tau, g) \in \mathbb{R}^2$$

It is assumed that $\mathbb{E}[T]$ and $\mathbb{E}[G]$ are finite. Since all random vectors (T_i, G_i) have the same distribution, it holds that

$$\mathbb{E}[T_i] = \mathbb{E}[T] \quad \forall i \in \{1, 2, 3, \ldots\}$$
$$\mathbb{E}[G_i] = \mathbb{E}[G] \quad \forall i \in \{1, 2, 3, \ldots\}$$

These renewal systems fit the requirements of Theorem 3. This is because

$$(\{(T_i, G_i)\}_{i=1}^{\infty} \text{ i.i.d.}) \implies (\{T_i\}_{i=1}^{\infty} \text{ i.i.d.} \quad \text{and} \quad \{G_i\}_{i=1}^{\infty} \text{ i.i.d.})$$

The converse implication does not always hold. Thus, Theorem 3 holds for renewal systems and also for some more general systems. However, renewal systems are the most common and the most practical applications of Theorem 3.

A. Examples

Example 1: (Bit arrivals) Suppose that packets arrive to a communication network with i.i.d. inter-arrival times $\{T_i\}_{i=1}^{\infty}$ that are exponentially distributed with rate parameter $\lambda > 0$, so that

$$f_T[t] = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbb{E}[T_i] = 1/\lambda$ seconds for all $i \in \{1, 2, 3, \ldots\}$. Suppose packet bit sizes are given by an i.i.d. sequence $\{B_i\}_{i=1}^{\infty}$ with $P[B_i = 1000] = P[B_i = 2000] = 1/2$. Compute the long term bit rate in units of bits/second.

Solution: Renewals occur at packet arrival times. Let $N(t)$ count the number of packets that arrive up to and including time t. By the renewal-reward theorem we have (almost surely):

$$\lim_{t\to\infty} \frac{\sum_{i=1}^{N(t)} B_i}{t} = \frac{\mathbb{E}[B]}{\mathbb{E}[T]} = \frac{1500}{1/\lambda} = 1500\lambda \text{ bits/sec}$$

and so the long term bit rate is $\alpha = 1500\lambda$ bits/sec.

Example 2: (Cloud processing) A cloud device performs back-to-back computational tasks. Each new task $i \in \{1, 2, 3, \ldots\}$ is independently a *type A task* with probability $1/4$ and a *type B task* with probability $3/4$. Type A tasks require 1 minute of processing and bring c_A dollars of revenue. Type B tasks independently require 1 minute of processing with probability $1/2$, and 2 minutes of processing with probability $1/2$, and bring c_B dollars of revenue. What is the long term profit rate α (in units of dollars/minute)?

Solution: Renewals occur upon each new task. Frames sizes are T_i and rewards are G_i:

$$T_i = \begin{cases} 1 & \text{with prob } (1/4) + (3/4)(1/2) \\ 2 & \text{with prob } (3/4)(1/2) \end{cases}$$

$$G_i = \begin{cases} c_A & \text{with prob } 1/4 \\ c_B & \text{with prob } 3/4 \end{cases}$$

Then with probability 1 we have

$$\alpha = \lim_{k\to\infty} \frac{\sum_{i=1}^k G_i}{\sum_{i=1}^k T_i} = \frac{\mathbb{E}[G_1]}{\mathbb{E}[T_1]} = \frac{(1/4)c_A + (3/4)c_B}{(5/8) + 2(3/8)} = \frac{2c_A + 6c_B}{11} \text{ dollars/min}$$

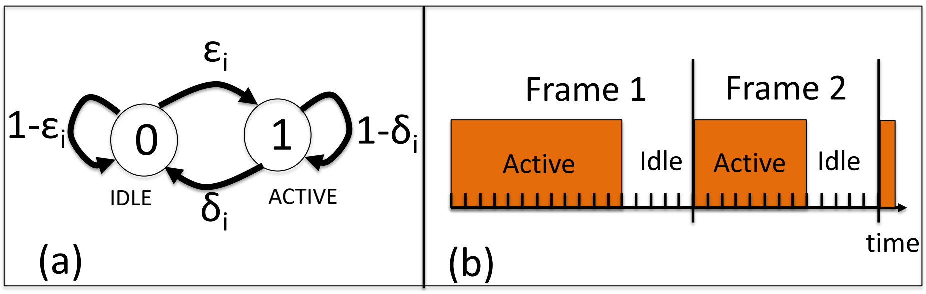

Fig. 3. (a) The 2-state DTMC for user $i \in \{1, \dots, n\}$; (b) The timeline of active and idle periods. Active and idle periods have average durations $1/\delta_i$ and $1/\epsilon_i$, respectively. In terms of *renewal theory*, an active period followed by an idle period can be viewed as a *renewal frame* and has *average frame size* $1/\delta_i + 1/\epsilon_i$.

Example 3: (Time averages for a DTMC) Consider a data network that operates in discrete time $t \in \{0, 1, 2, \dots\}$. The system has multiple users $i \in \{1, \dots, n\}$ that send data according to independent idle/active processes that are modeled by a 2-state DTMC shown in Fig. 3. When a user is in the active state on slot t, it sends one new packet into the network. When a user is idle on slot t it sends nothing. Each user $i \in \{1, ..., n\}$ has transition probabilities ϵ_i and δ_i as shown in the figure. For simplicity, this problem treats only user, so we use transition probabilities $\epsilon \in (0, 1)$ and $\delta \in (0, 1)$ (for simplicity we suppress the i subscript). Each renewal frame is defined by an active period followed by an idle period, as shown in Fig 3b. For each frame $i \in \{1, 2, 3, \dots\}$ define the frame size $T_i = B_i + I_i$, where B_i is the duration of the busy period of frame i and I_i is the duration of the idle period of frame i. Define the reward $G_i = B_i$. The reward rate is then the total fraction of time the system is busy:

$$\text{fraction of time busy} = \lim_{k\to\infty} \left[\frac{\text{Total time busy during } k \text{ frames}}{\text{Total time of } k \text{ frames}}\right] = \lim_{k\to\infty} \frac{\sum_{i=1}^k G_i}{\sum_{i=1}^k T_i}$$

Compute the fraction of time busy using renewal-reward theory and compare to the steady state probability of being busy.

Solution: State 1 is the active state. We have with prob 1

$$\text{fraction of time busy} = \frac{\mathbb{E}[G_1]}{\mathbb{E}[T_1]} = \frac{1/\delta}{1/\delta + 1/\epsilon} = \frac{\epsilon}{\epsilon + \delta}$$

which is the same as the steady state probability of being in the active state (state 1).

Example 4: (Ratio of expectations) Consider a renewal system with i.i.d. vectors $\{(T_i, G_i)\}_{i=1}^\infty$ given as follows: Let $\{T_i\}_{i=1}^\infty$ be i.i.d. $Unif[1, 9]$. Let $\{V_i\}_{i=1}^\infty$ be i.i.d. $Unif[0, 1]$ and assume the processes $\{T_i\}_{i=1}^\infty$ and $\{V_i\}_{i=1}^\infty$ are independent. Define $G_i = T_i^3 + V_i$ for $i \in \{1, 2, 3, \dots\}$. Define $T = T_1$ and $G = G_1$. What is the correct value of long term average reward $\alpha = \lim_{t\to\infty} R(t)/t$?

a) $\alpha = \mathbb{E}\left[\frac{G}{T}\right] = \mathbb{E}\left[T^2\right] + \mathbb{E}[V]\,\mathbb{E}\left[\frac{1}{T}\right] = \frac{91}{3} + (0.5)(\frac{\log(3)}{4}) = 30.471$

b) $\alpha = \frac{\mathbb{E}[G]}{\mathbb{E}[T]} = \frac{205+0.5}{5} = 41.1$

Solution: The answer is (b) (by the renewal-reward theorem). While (a) and (b) give correct numerical values for $\mathbb{E}[G/T]$ and $\mathbb{E}[G]/\mathbb{E}[T]$, respectively, the value $\mathbb{E}[G/T]$ is irrelevant. In particular, the correct answer is the *ratio of expectations* not the *expectation of the ratio*. Thus, the answer only depends on the *marginal distributions* for T and G separately, rather than the joint distribution for the random vector (T, G). Students who are not convinced are encouraged to simulate this example on a computer over n iterations, where n is a large integer (say, $n = 10^5$ or $n = 10^6$), and compare the resulting values of $\frac{\sum_{i=1}^n G_i}{\sum_{i=1}^n T_i}$ and $\frac{1}{n}\sum_{i=1}^n (G_i/T_i)$.

B. Discrete time

So far we have treated *continuous time systems* that can be evaluated at any real number time $t \geq 0$. Many systems are designed to operate in slotted time with time steps that take integer values $t \in \{0, 1, 2, 3, \ldots\}$. Such systems are called *discrete time systems*. Renewal systems in discrete time can be viewed as a special case of continuous time renewal systems where $\{(T_i, G_i)\}_{i=1}^{\infty}$ are i.i.d. vectors, but where each random variable T_i is assumed to take a positive integer value. For example, if we have

$$T_i = \begin{cases} 1 & \text{with prob } 1/2 \\ 2 & \text{with prob } 1/4 \\ 3 & \text{with prob } 1/4 \end{cases}$$

then each new renewal starts at an integer time $T_1, T_1 + T_2, T_1 + T_2 + T_3$, and so on. The time average rate of rewards satisfies

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[G]}{1(1/2) + 2(1/4) + 3(1/4)} = (4/7)\mathbb{E}[G] \quad \text{almost surely}$$

When viewed in continuous time, the functions $R(t)$ and $N(t)$ are assumed to hold their values fixed at the value of the latest integer time. The above limit can also be viewed as taking $t \to \infty$ over the set of positive integers.

A special case of discrete time renewal systems is a system with 1-slot frame durations, so that $T_i = 1$ surely for all $i \in \{1, 2, 3, \ldots\}$. In this case, rewards come i.i.d. every slot. For these systems, the rewards are often re-indexed as $G(0), G(1), G(2)$ and so on, instead of G_1, G_2, G_3. Thus, the first slot is slot 0.

C. Extensions of renewal theory

Appendix B of these notes give the following extensions:

- Delayed renewal-reward theorem: The same limiting behavior holds when the random variables T_1 and G_1 associated with frame 1 can deviate from the the i.i.d. behavior of $\{T_i\}_{i=2}^{\infty}$ and $\{G_i\}_{i=2}^{\infty}$.
- Renewal-reward with expectations: When the assumptions of Theorem 3 hold, can we also say:

$$\lim_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} = \frac{1}{\mathbb{E}[T]} \quad ?$$
$$\lim_{t\to\infty} \frac{\mathbb{E}[R(t)]}{t} = \frac{\mathbb{E}[G]}{\mathbb{E}[T]} \quad ?$$

  [The answer is "yes" but it is not easy to prove. The first limit is called the "Elementary renewal theorem" but its proof is not elementary. The second limit is called the "Elementary renewal-reward theorem."][6]
- Reward sprinkling: Instead of assuming the full reward G_i arrives in bulk at the end of frame i, what happens if different pieces of G_i are sprinkled over the duration of frame i? [The answer is subtle, as described in Appendix B. Fortunately, for most practical systems of interest the answer for sprinkled rewards is the same as for nonsprinkled rewards.]

V. BASIC MAC MODEL: n USERS, EACH WITH INFINITE PACKETS TO SEND

Fix n as a positive integer. Suppose there are n network users that want to send data to a common access point. Suppose the timeline is segmented into fixed length *time slots*. The time slot length is typically a fraction of a second in duration. All users know the slot duration and are synchronized in time so that they agree when a new time slot begins. Each user has an infinite amount of data to send. Assume the data of each user is packaged into an infinite sequence of *fixed length packets*. The time slot duration is assumed to be sufficient for exactly one user to send exactly one packet.

Time slots are indexed by $t \in \{0, 1, 2, \ldots\}$. At the start of a new time slot t, each user decides to send either one packet or no packet. A packet that is sent by a particular user on slot t is successfully received at the access point if and only if there are no other users that attempt to send packets on that same slot. If two or more users attempt to send packets on the same slot then a *collision* occurs and no information is delivered to the access point. If nobody attempts to send a packet then an *idle* occurs and again no information is delivered. Each user is assumed to know the success/idle/collision outcome at the end of every slot. That is, it is assumed that each user receives *0/1/C feedback* at the end of every slot:

- Idle [0]: No users sent a packet.
- Success [1]: Exactly one user sent a packet.
- Collision (C): Two or more users sent a packet.

Each user has some algorithm for making transmission decisions over time. This section focuses on the ideal scenario where the feedback is perfectly known by all users by the end of each slot.

[6]Notice that we do not say $\lim_{t\to\infty} \mathbb{E}[N(t)]/t = 1/\mathbb{E}[T]$ holds "almost surely" because $\mathbb{E}[N(t)]/t$ is a deterministic function of t, so its limit is in the usual sense of limits of real-valued functions as defined in basic calculus courses.

A. Application scenarios, hidden terminals, and further questions

1) Wireless network: Suppose the n users each send data to the access point via their individual wireless devices. The devices are synchronized according to common time slots. The Idle/Success/Collision feedback can be determined either by having each user perform *carrier sensing* to determine channel activity (which suffers from a *hidden terminal* problem if some users cannot hear transmissions of their peers), or by having the receiver explicitly send ACK/NACK information over a control channel (which helps to combat hidden terminals while possibly creating an inefficiency if the control channel uses the same spectrum resources as the data transmissions).

Questions: How can the devices maintain this time synchronization? What happens if there are timing offsets at each device? How can each device receive the assumed 0/1/C feedback? Does the access point explicitly send this feedback, or can the devices determine this feedback themselves by sensing the channel? Do these questions have technology-dependent answers? What if there are *hidden terminals* so that some devices cannot hear when others are transmitting?

2) Shared cable network: Suppose the n users are computers that want to send data over a shared cable, sometimes called a shared bus. Only one computer can use the cable at a time, else, the result is a garbled signal. This is a classical model for "old-school" ethernet.[7]

As in the wireless case, the same questions of synchronization and feedback apply. Fortunately, the hidden terminal problem disappears in this context: When one computer is transmitting, the signal propagates over the shared cable so that all other devices (eventually) hear the transmission. The worst-case time required for a signal initiated by one computer to be detected by another can be computed by calculating the time required for a signal to travel from one side of the cable to the other while moving at the speed of light.

B. Throughput and fairness

For each $i \in \{1, \ldots, n\}$ and each $t \in \{0, 1, 2, \ldots\}$ define

$$\begin{aligned} X_i[t] &= \begin{cases} 1 & \text{if user } i \text{ successfully sends a packet on slot } t \\ 0 & \text{otherwise} \end{cases} \\ X[t] &= \sum_{i=1}^{n} X_i[t] \end{aligned} \tag{22}$$

Notice that at most one user can be successful per slot and so for each $i \in \{1, \ldots, n\}$ we have

$$X_i[t] = 1 \implies X_j[t] = 0 \quad \forall j \neq i$$

Thus

$$X[t] = \begin{cases} 1 & \text{if some user was successful on slot } t \\ 0 & \text{if nobody was successful on slot } t \end{cases}$$

For each positive integer T, define $\overline{X}_i[T]$ and $\overline{X}[T]$ as the throughput for user i and the total throughput, respectively, over the first T slots:

$$\begin{aligned} \overline{X}_i[T] &= \frac{1}{T}\sum_{t=0}^{T-1} X_i[t] \quad (packets/slot) \quad \forall i \in \{1, \ldots, n\} \\ \overline{X}[T] &= \frac{1}{T}\sum_{t=0}^{T-1} X[t] \quad (packets/slot) \end{aligned}$$

Since $X_i[t]$ and $X[t]$ are either 0 or 1 on each slot t, it is clear that

$$\begin{aligned} &0 \leq \overline{X}_i[T] \leq 1 \quad \forall i \in \{1, \ldots, n\} \\ &0 \leq \overline{X}[T] \leq 1 \end{aligned}$$

Define $\overline{X}_i$ and $\overline{X}$ as the limiting (infinite horizon) throughput:

$$\begin{aligned} \overline{X}_i &= \lim_{T\to\infty} \tfrac{1}{T}\textstyle\sum_{t=0}^{T-1} X_i[t] \quad (packets/slot) \quad \forall i \in \{1, \ldots, N\} \\ \overline{X} &= \lim_{T\to\infty} \tfrac{1}{T}\textstyle\sum_{t=0}^{T-1} X[t] \quad (packets/slot) \end{aligned}$$

where the limits are assumed to exist (with probability 1) for simplicity. Assuming these infinite horizon limits exist, it is clear that $0 \leq \overline{X}_i \leq 1$ and $0 \leq \overline{X} \leq 1$.

It is desirable to design algorithms for each user to ensure that the total throughput $\overline{X}$ is as large as possible, meaning that it is as close to 1 as possible. One potential method is to have user 1 transmit on every slot $t \in \{0, 1, 2, \ldots\}$ and have

[7] Modern ethernet is organized differently using packet switches. Switches and related topics are considered later in the course.

users $\{2, 3, ..., n\}$ never transmit. This ensures that there are never any idle slots or collision slots, so that $X[t] = 1$ for all $t \in \{0, 1, 2, \ldots\}$ and $\overline{X} = 1$. There are two problems with this:

1) It is not clear how to determine which of the N users is user 1. Indeed, if there are n people in a room, how do we determine which is person number 1? Also, how do we tell everyone that user 1 is the only one allowed to transmit?
2) While this strategy maximizes throughput ($\overline{X} = 1$), it is *unfair* to users $\{2, 3, \ldots, n\}$.

To address the second point, it is useful to have a quantifiable measure of fairness. We want fairness to be quantified in terms of the vector of throughputs $(\overline{X}_1, \ldots, \overline{X}_n)$, not just the sum throughput. Intuitively, we want the components of this vector to be close to each other so that every user receives a similar throughput value. One standard measure of fairness is the *Jain fairness index*:

$$F(\overline{X}_1, \ldots, \overline{X}_n) = \frac{(\sum_{i=1}^n \overline{X}_i)^2}{n \sum_{i=1}^n \overline{X}_i^2}$$

It is implicitly assumed that the vector $(\overline{X}_1, \ldots, \overline{X}_n)$ has at least one positive component (else the fairness index F reduces to the undefined expression 0/0). As long as the throughput vector consists of nonnegative values and has at least one positive component, it can be shown that

$$0 < F(\overline{X}_1, \ldots, \overline{X}_n) \leq 1$$

A large fairness index (close to 1) is desirable because it means that the components of vector $(\overline{X}_1, \ldots, \overline{X}_n)$ are close together. This is illuminated in the following lemma.

Lemma 1: Fix N as a positive integer and define

$$D_n = \left\{(x_1, \ldots, x_n) : x_i \geq 0 \quad \forall i \in \{1, \ldots, n\}, \sum_{i=1}^n x_i > 0\right\}$$

Define $F : D_n \to \mathbb{R}$ by

$$F(x_1, \ldots, x_n) = \frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2}$$

Then

a) $0 < F(x_1, \ldots, x_n) \leq 1$ for all $(x_1, \ldots, x_n) \in D_n$.
b) $F(x_1, \ldots, x_n) = 1$ if and only if $x_1 = x_2 = ... = x_n$.

Proof: Fix $(x_1, \ldots, x_n) \in D_n$. Let I be a random variable that is uniform over the index set $\{1, \ldots, n\}$, that is, $P[I = i] = 1/n$ for all $i \in \{1, \ldots, n\}$. Define the random variable $X = x_I$. The random variable X is bounded because $0 \leq X \leq \max\{x_1, \ldots, x_n\}$ and hence it has a finite mean and variance. By the law of total expectation we have

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X|I = i] (1/n) = \frac{1}{n}\sum_{i=1}^n x_i$$

$$\mathbb{E}\left[X^2\right] = \sum_{i=1}^n \mathbb{E}\left[X^2|I = i\right] (1/n) = \frac{1}{n}\sum_{i=1}^n x_i^2$$

and so

$$F(x_1, ..., x_n) = \frac{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}{\frac{1}{n}\sum_{i=1}^n x_i^2} = \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \tag{23}$$

Since $Var(X) \geq 0$ we know $\mathbb{E}\left[X^2\right] \geq \mathbb{E}[X]^2$ and so $F(x_1, \ldots, x_n) \leq 1$. This proves part (a).

To prove part (b), first observe that if $x_1 = x_2 = ... = x_n$ then $F(x_1, \ldots, x_n) = 1$. Now suppose $F(x_1, \ldots, x_1) = 1$. From (23) we obtain

$$1 = \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}$$

which means $\mathbb{E}\left[X^2\right] = \mathbb{E}[X]^2$, so $Var(X) = 0$. However, $Var(X) = 0$ can only occur if X takes a constant value with probability 1, which is only possible if $x_1 = x_2 = ... = x_n$. □

Observe that if $\overline{X}_i = \frac{1}{n10^{100}}$ for all $i \in \{1, \ldots, n\}$ then fairness index satisfies:

$$F(\overline{X}_1, \ldots, \overline{X}_n) = 1$$

while the throughput is extremely low:

$$\overline{X} = \sum_{i=1}^n \overline{X}_i = \frac{1}{10^{100}}$$

Therefore, it is not enough to maximize the fairness alone. Also, it is not enough to maximize the throughput alone. We want *both* the fairness and the throughput to be large. This is called a *bicriteria optimization problem* because we want to maximize two different performance objectives (fairness and throughput). In most optimization scenarios, it is impossible to maximize two different performance objectives at the same time. Thus, we must settle for a solution that obtains a desirable *tradeoff* between the two objectives. However, in this case, we observe that if $(\overline{X}_1, \ldots, \overline{X}_n) = (1/n, 1/n, \ldots, 1/n)$ then both throughput and fairness are maximized:

$$F(1/n, 1/n, ..., 1/n) = 1, \quad \sum_{i=1}^{n}(1/n) = 1$$

Therefore, it is desirable to develop algorithms that yield a throughput vector that is close to the vector $(1/n, 1/n, \ldots, 1/n)$.

C. Round-robin

An easy way to achieve a throughput vector of $(1/n, 1/n, \ldots, 1/n)$ is the following *round-robin* scheme:

- User 1 transmits at time 0.
- User 2 transmits at time 1.
- ...
- User n transmits at time $n-1$.
- Repeat.

In other words, user $i \in \{1, \ldots, n\}$ transmits at time slot t if and only if $t \mod n = i - 1$. This ensures there is one and only one user that transmits on each slot. Thus, there are no collisions and every slot has a successful packet transmission. This scheme is desirable and is often implemented (when possible). However, it requires some preliminary coordination. How can we know the number of users? How do the users know their user index $i \in \{1, \ldots, n\}$? What happens if the number of users changes over time?

D. Basic slotted Aloha with a common transmit probability p

Assume $n \geq 2$, so that there is more than one user. The basic *slotted Aloha* protocol is for each user to independently transmit with probability p on every slot $t \in \{0, 1, 2, \ldots\}$. Let $Y[t]$ denote the number of transmitters on a given slot t. Then $Y[t] \in \{0, 1, \ldots, n\}$ and $Y[t] \sim \text{Binomial}(n, p)$ for each slot t, so

$$P[Y[t] = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad \forall k \in \{1, 2, ..., n\} \tag{24}$$

$$\mathbb{E}[Y(t)] = np \tag{25}$$

Since the same procedure is independently repeated every slot, the random variables $\{Y[t]\}_{t=0}^{\infty}$ are independent and identically distributed (i.i.d.). Slot t yields an *idle* if $Y[t] = 0$; a *success* if $Y[t] = 1$; a *collision* if $Y[t] \geq 2$. In particular, since $X[t]$ is a Bernoulli variable that is 1 if and only if there is a success on slot t, we have for all $t \in \{0, 1, 2, \ldots\}$:

$$\text{Success on slot } t \iff U[t] = 1 \iff X[t] = 1$$

$$\begin{aligned} P[X[t] = 1] &= np(1-p)^{n-1} \\ P[X[t] = 0] &= 1 - np(1-p)^{n-1} \\ \mathbb{E}[X[t]] &= np(1-p)^{n-1} \end{aligned} \tag{26}$$

Since $\{X[t]\}_{t=0}^{\infty}$ are i.i.d. Bernoulli random variables, we have by the law of large numbers

$$\boxed{\overline{X} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} X[t] = np(1-p)^{n-1} \quad \text{(with prob 1)}} \tag{27}$$

We also observe that for each user $i \in \{1, \ldots, n\}$, the sequence $\{X_i[t]\}_{t=0}^{\infty}$ is i.i.d. Bernoulli and so

$$\overline{X}_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} X_i[t] = \mathbb{E}[X_i[0]] \quad \text{(with prob 1)}$$

By symmetry, we have

$$\mathbb{E}[X_1[0]] = \mathbb{E}[X_2[0]] = ... = \mathbb{E}[X_n[0]]$$

and so

$$\mathbb{E}[X[0]] = \sum_{j=1}^{n} \mathbb{E}[X_j[0]] = n\mathbb{E}[X_i[0]] \quad \forall i \in \{1, \ldots, n\}$$

Thus, $\mathbb{E}[X_i[0]] = \frac{\mathbb{E}[X[0]]}{n} = p(1-p)^{n-1}$ (which uses (26)). That is,

$$\boxed{\overline{X}_i = p(1-p)^{n-1} \quad \forall i \in \{1, \ldots, n\} \quad \text{(with prob 1)}} \tag{28}$$

E. Optimizing p for slotted Aloha

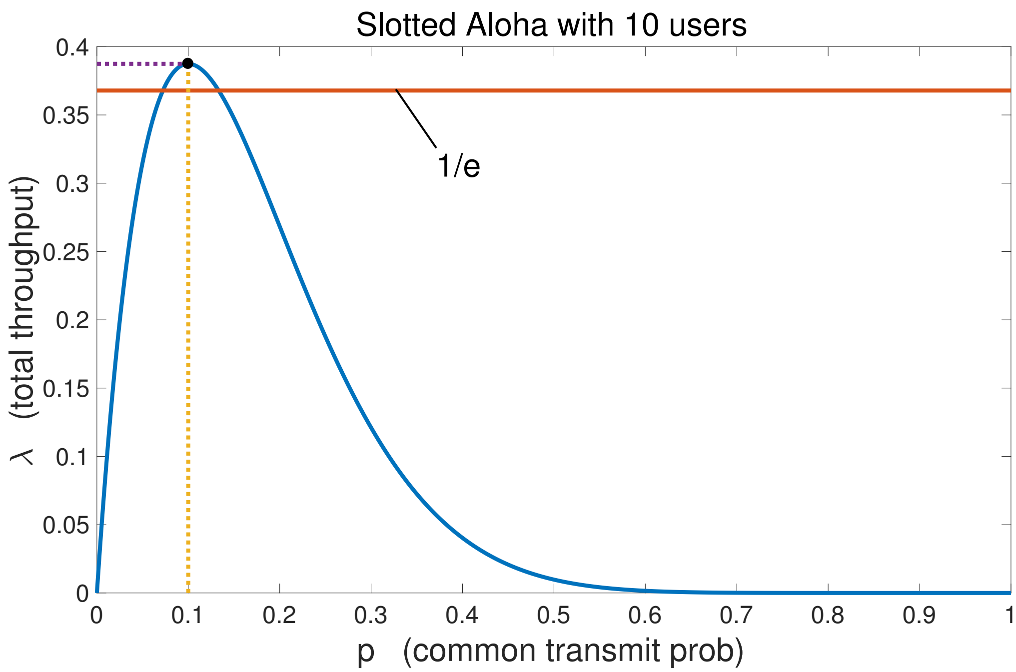

Fig. 4. A plot of throughput versus p for slotted Aloha with 10 users. Note that $p^* = 1/10$, $\lambda^* = (1-1/10)^9 \approx 0.3874$.

For any $p > 0$ the throughput $\overline{X}_i$ for user i, given in (28), is positive and is the same for all $i \in \{1, \ldots, n\}$. Hence, slotted Aloha yields a Jain fairness index of 1 for any commonly used transmit probability $p > 0$:

$$F(\overline{X}_1, \ldots, \overline{X}_n) = 1$$

However, the value of total throughput $\overline{X} = np(1-p)^{n-1}$ in (27) varies with p. Define λ as this total throughput value:

$$\lambda = np(1-p)^{n-1}$$

Observe that if $p = 0$ or $p = 1$ then $\lambda = 0$. This is intuitively clear because if $p = 0$ then nobody ever transmits and every slot is idle, while of $p = 1$ then everyone always transmits and so every slot is a collision (recall that $n \geq 2$). A plot of the curve $np(1-p)^{n-1}$ is shown for the case $n = 10$ users in Fig. 4. The value of $p \in [0, 1]$ can be optimized to give the largest throughput for this particular scheme:

$$\frac{d(np(1-p)^{n-1})}{dp} = 0 \implies n(1-p)^{n-1} - np(n-1)(1-p)^{n-2} = 0$$
$$\implies \boxed{p^* = 1/n}$$

Since the expected number of transmitters on a given slot is np (recall (25)), it is interesting that, at the optimal transmission probability $p^* = 1/n$, the expected number of transmitters is 1.

Let $p_n^* = 1/n$ denote this optimal value as a function of the number of users $n \in \{2, 3, 4, \ldots\}$. For n users, this optimized probability p_n^* yields a total throughput λ_n^* of

$$\boxed{\lambda_n^* = (1 - 1/n)^{n-1}}$$

A plot of λ_n^* versus n is shown in Fig. 5. It is insightful to observe how p_n^* and λ_n^* behave as $n \to \infty$:

$$\lim_{n\to\infty} p_n^* = \lim_{n\to\infty} 1/n = 0$$
$$\lim_{n\to\infty} \lambda_n^* = \lim_{n\to\infty} (1-1/n)^{n-1} = 1/e \approx 0.367879$$

The latter limit can be computed by writing $(1-1/n)^{n-1} = (1-1/n)^n(1-1/n)^{-1}$ and using the following basic fact from calculus (for the case $x = -1$):

Fact 1: (A basic limit)

$$\lim_{m\rightarrow\infty}\left(1+\frac{x}{m}\right)^m = e^x \quad \forall x \in \mathbb{R}$$

Proof: Fix $x \in \mathbb{R}$. Define $y_m = (1+x/m)^m$ and observe that $y_m > 0$ for all sufficiently large m, so that $\log(y_m)$ is well defined for all sufficiently large m. Then

$$\begin{aligned}
\lim_{m\rightarrow\infty}\log(y_m) &= \lim_{m\rightarrow\infty} m\log(1+x/m) \\
&= \lim_{m\rightarrow\infty}\frac{\log(1+x/m)}{1/m} \\
&\overset{(a)}{=} \lim_{m\rightarrow\infty}\frac{\frac{-x/m^2}{1+x/m}}{-1/m^2} \\
&= x
\end{aligned}$$

where (a) holds by L'Hopital's rule. It follows that $y_m = e^{\log(y_m)} \rightarrow e^x$. □

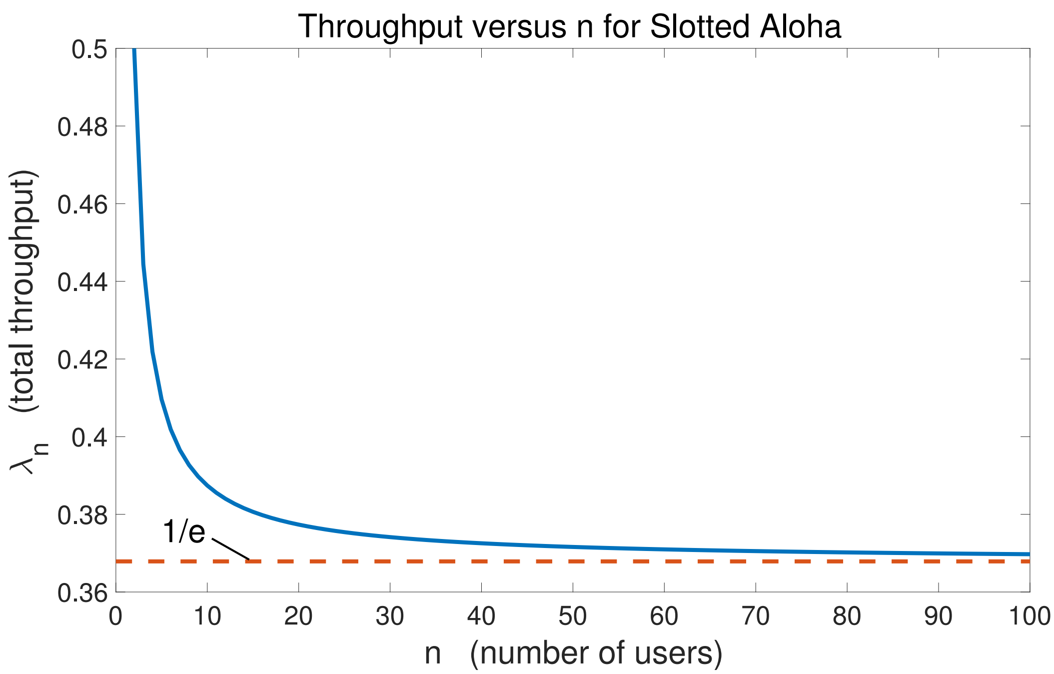

Fig. 5. Optimized throughput $\lambda_n^* = (1-1/n)^{n-1}$ versus number of users n for slotted Aloha.

F. Slotted Aloha with capture

Consider a slotted Aloha system with $n \geq 2$ users and with 0/1/C feedback, but with the difference that once a packet is successful, it transmits $k-1$ additional times in a row, where k is a parameter that is specified in advance. All other users are silent for these $k-1$ transmissions. In particular, once a user is successful once, that use *captures the channel* and sends a total burst of k packets in a row, where the last $k-1$ packets suffer no contention with other users. This can be viewed as a *renewal system* where renewal frames last for either 1 slot or k slots. Each user indepdendently transmits with probability p at the start of each renewal and receives 0/1/C feedback at the end of that slot. We have three possibilities:

- Idle: No users transmit at the initial renewal slot. Renewal frames last one slot.
- Collision: Two or more users transmit at the initial renewal slot. Renewal frames last one slot.
- Success: Exactly one user transmits at the initial renewal slot. That use captures the channel for the next $k-1$ slots. Renewal frames last k slots and there are k successes.

Let $T_i \in \{1, k\}$ be the size of each renewal frame i and let G_i be the total number of successes on renewal frame i. The total system throughput λ is

$$\lambda = \lim_{k\rightarrow\infty}\frac{\sum_{i=1}^k G_i}{\sum_{i=1}^k T_i} = \frac{\mathbb{E}[G_1]}{\mathbb{E}[T_1]} \quad \text{packets/slot (with prob 1)}$$

It remains only to compute the expectations $\mathbb{E}[G_1]$ and $\mathbb{E}[T_1]$ associated with frame 1:

$$\mathbb{E}[G_1] = kP[Success] = knp(1-p)^{n-1}$$
$$\mathbb{E}[T_1] = kP[Success] + 1(1 - P[Success]) = knp(1-p)^{n-1} + 1 - np(1-p)^{n-1}$$

Hence

$$\boxed{\lambda_n = \frac{knp(1-p)^{n-1}}{(k-1)np(1-p)^{n-1} + 1}}$$

By system symmetry the individual throughput of each user $i \in \{1, \ldots n\}$ is λ_n/n.

Remarkably, for any given integers $n \geq 2$, $k \geq 1$, the expression for λ in the above boxed formula is again maximized over all $p \in [0,1]$ when $p^* = 1/n$.[8] This yields throughput:

$$\lambda_n^* = \frac{k(1-1/n)^{n-1}}{(k-1)(1-1/n)^{n-1} + 1}$$

Taking $n \to \infty$ gives

$$\lim_{n\to\infty} \lambda_n^* = \frac{k}{k-1+e}$$

When $k = 1$ this reduces to the standard slotted Aloha throughput of $1/e$. However, using a larger value of k results in a significantly larger throughput. Throughput gains for various k are shown in the table below (Fig. 6). The table shows a significant gain by moving from $k = 1$ to $k = 2$ to $k = 3$. The table also shows that when $k = 10^6$ the throughput is close enough to 1 for all practical purposes. However, the value of k also affects *system delay*: If k is too large then a single lucky user captures the channel for, say, an hour, while all other users wait! In practice, using $k \in \{2, ..., 5\}$ is reasonable because users are happy to wait a few extra slots to take advantage of their share of the increased throughput.

k	$\lambda^* = k/(k-1+e)$
1	$1/e = 0.3679$
2	0.5379
3	0.6358
4	0.6995
5	0.7442
10	0.8534
100	0.9831
10^6	0.999998

Fig. 6. Throughput of slotted Aloha with k-slot capture for various k values and large n (that is, $n \to \infty$).

Theoretically, it is interesting to note that for any fixed $n \geq 2$ and any fixed p such that $0 < p < 1$, throughput can still be pushed to 1 by making k large:

$$\lim_{k\to\infty} \frac{knp(1-p)^{n-1}}{(k-1)np(1-p)^{n-1} + 1} = 1$$

Overall, the capture concept allows total system throughput to increase. There are two disadvantages: First, delay can be large when k is large. Second, the algorithm requires k to be known, and requires all users to closely monitor system behavior so they can be aware of the start of new frames. This is difficult if *new users* arrive because they may not know where frames start! One way to handle this is to have the access point send beacon signals at the start of every new frame. Another way is for new users to silently listen to the system until they hear an idle or collision followed by k successes in a row.

G. ZigZag decoding

It is shown in [4] that it is often possible to reliably detect the *number* of colliding packets even if the data from those packets cannot be retrieved. The ZigZag decoding idea, developed in [4], is a method that can decode two colliding packets in exactly two time slots. Consider a situation where exactly two packets collide: The ZigZag feature automatically tells these same two users to transmit the same packets *again* on the next slot (while all other users are silent). This surely produces another collision on the next slot. However, random timing offsets between the transmissions, along with signal processing techniques, enables the complete decoding of both packets after unraveling their entangled collisions from both slots. Implementation of this idea was shown in [4] to significantly improve performance in multiple access systems. A soft-decision version called

[8] This is easily seen by assuming $0 < p < 1$ and writing $\lambda_n^* = \frac{k}{(k-1) + \frac{1}{np(1-p)^{n-1}}}$, which is maximized when $np(1-p)^{n-1}$ is maximized.

SigSag that improves error rates is in [5]. The analysis in this subsection is inspired by the ZigZag MAC analysis in [7] (see also MAC protocols for other multi-packet reception scenarios in [6]).

Consider the following frame-based analysis in a multi-access system that uses ZigZag. Assume there are n users, all who want to send an infinite number of fixed-length packets. The system operates over frames of size either 1 slot or 2 slots. Let $\{(T_k, G_k)\}_{k=1}^{\infty}$ be the process of rewards and frame durations. The value T_k is the size of frame k and the value G_k is the number of successful packets sent to the receiver on frame k. At the start of each frame k, all users independently transmit with probability p. Let N_k be the number of users who transmit at the start of frame k. Then

- If $N_k = 0$ then frame k is an idle and $(T_k, G_k) = (1, 0)$.
- If $N_k = 1$ then frame k is a success and $(T_k, G_k) = (1, 1)$.
- If $N_k = 2$ then frame k is a ZigZag and $(T_k, G_k) = (2, 2)$.
- If $N_k \geq 3$ then frame k is a collision and $(T_k, G_k) = (1, 0)$.

What is the long term total throughput λ (in units of packets/slot)? By renewal-reward theory we have

$$\boxed{\lambda_n = \frac{\mathbb{E}\left[G\right]}{\mathbb{E}\left[T\right]} = \frac{np(1-p)^{n-1} + 2\binom{n}{2}p^2(1-p)^{n-2}}{1 + \binom{n}{2}p^2(1-p)^{n-2}}}$$

To consider throughput for large-n, fix $\theta > 0$ and define $p = \theta/n$ (assuming $n \geq \theta$). Then

$$\lambda = \lim_{n\to\infty} \lambda_n = \frac{\theta e^{-\theta} + \theta^2 e^{-\theta}}{1 + \frac{\theta^2}{2}e^{-\theta}}$$

A plot of λ versus θ is given in Fig. 7. This is optimized at

$$\theta^* = 1.49951$$
$$\lambda^* = 0.66884$$

Thus, the optimal transmission probability for each user is θ^*/n. Since $\theta^* > 1$, transmissions are more aggressive than those for slotted Aloha (recall that $\theta = np$ is the expected number of transmitters at the start of a frame). This is intuitive because collisions in this ZigZag scenario are less costly. In particular, collisions of exactly two packets are good! As expected, the throughput of 0.66884 is significantly larger than the $1/e = 0.3679$ throughput of slotted Aloha. A more detailed ZigZag analysis incorporates the probability $\epsilon > 0$ that a ZigZag frame fails to decode. The only change in the analysis is that the expected reward in each frame is decreased by an amount $O(\epsilon)$.

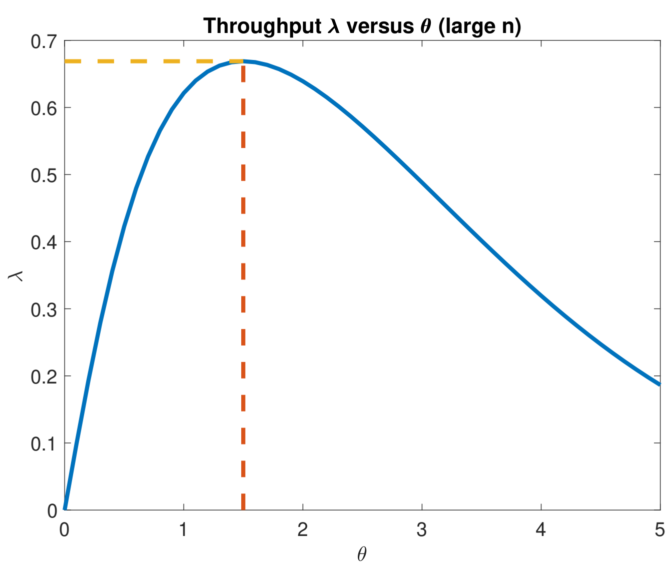

Fig. 7. Throughput of ZigZag versus θ (for large n). At optimality we have $\theta^* = 1.49951$ and $\lambda^* = 0.66884$.

H. Carrier Sense Multiple Access (CSMA)

This subsection presents a refinement on slotted Aloha called (slotted) Carrier Sense Multiple Access (CSMA). Again assume there are a fixed number of users n and that each user has an infinite number of packets to send. The idea is to assume that all users can sense the channel to detect idle slots. The sensing time is assumed to be significantly smaller than a timeslot.

Thus, if all users detect that a slot will be idle, that slot can be cut short to reduce time waste. Minimally, the sensing time should be at least as large as the time required for a signal of one user to propagate to the location of another user. Thus, in practice, the worst-case sensing time is sized according to an assumed worst-case distance between two users.

Suppose the time to transmit a packet is 1 timeslot. Suppose the worst-case time to sense the channel is $\beta = 1/m$ for some positive integer $m \geq 2$, so that $0 < \beta < 1$. The time unit β is called a "mini-slot." This time β accounts for misalignment of time clocks at different users as well as propagation delay of a signal. To understand the savings, we first compare slotted Aloha with slotted CSMA in a simple sample path example. Suppose the first 7 events that occur in slotted Aloha are $C, 0, 0, 0, 1, C, 1$. This takes a total of 7 time units and delivers exactly 2 packets. The same 7 events still occur in slotted CSMA, but now the idle slots are reduced to β units of time. The total time taken in a CSMA implementation is $4+3\beta < 7$, and the same number of successes have occurred in this (smaller) period of time.

This gives intuition. However, to successfully implement CSMA we will need to pad the start of each transmission event with an idle slot of size β to allow users time to detect the end of the previous event. Since there were four transmission events (including two collisions and two successes) the padding will increase these slots by β for each (for a total of 4β more time units) so the total time under CSMA is $4 + 7\beta$. If β is significantly smaller than 1, then this value is still smaller than the 7 units of time required for slotted Aloha.

System events now operate at the start of mini-slots of size β. There are variable-sized frames. The start of each renewal frame is an idle mini-slot of size β where users sense the channel as idle (to detect the end of the previous frame). During this time users also decide whether or not to transmit. If they decide to transmit, that transmission will start at the end of this mini-slot and will last 1 unit of time (m mini-slots). The users get 0/1/C feedback, but feedback about idles [0] is obtained within β units of time and the size and reward of each renewal frame is determined accordingly. The (size, reward) vector (T_i, G_i) for renewal frame i is determined by the feedback:

- Success: Renewal size is $\beta + 1$ time units, 1 packet is delivered. $(T_i, G_i) = (\beta + 1, 1)$.
- Collision: Renewal size is $\beta + 1$ time units, 0 packets are delivered. $(T_i, G_i) = (\beta + 1, 0)$.
- Idle: Renewal size is β time units, 0 packets are delivered. $(T_i, G_i) = (\beta, 0)$.

We also have:

$$\begin{aligned} P[Success] &= np(1-p)^{n-1} \\ P[Idle] &= (1-p)^n \\ P[Collision] &= 1 - (1-p)^n - np(1-p)^{n-1} \end{aligned}$$

Let $\lambda = \lim_{k\to\infty} \frac{\sum_{i=1}^k G_i}{\sum_{i=1}^k T_i}$ be the total throughput in units of packets/time. By renewal-reward theory we have

$$\lambda = \frac{\mathbb{E}[G_1]}{\mathbb{E}[T_1]} \quad \text{(with prob 1)}$$

It remains to calculate $\mathbb{E}[G_1]$ and $\mathbb{E}[T_1]$:

$$\begin{aligned} \mathbb{E}[G_1] &= P[Success] = np(1-p)^{n-1} \\ \mathbb{E}[T_1] &= \beta(1-p)^n + (1+\beta)(1-(1-p)^n) = \beta + 1 - (1-p)^n \end{aligned}$$

Thus,

$$\boxed{\lambda = \frac{np(1-p)^{n-1}}{\beta + 1 - (1-p)^n}}$$

For given β and n values, the probability p in the above expression can be optimized. For better intuition, suppose that we use:

$$p = \theta/n$$

for some $\theta \geq 0$ (again, $\theta = np$ is the expected number of transmitters at the start of a frame). We know that for usual slotted Aloha, the optimal transmit probability is $1/n$, and so assuming a probability that is proportional to $1/n$ (with proportionality constant θ) is reasonable. Substituting $p = \theta/n$ into the above boxed expression gives:

$$\lambda_{\theta,n} = \frac{\theta(1-\theta/n)^{n-1}}{\beta + 1 - (1-\theta/n)^n}$$

where we use subscript (θ, n) to emphasize dependence on θ and n. Now suppose θ is held fixed while $n \to \infty$:

$$\lim_{n\to\infty} \lambda_{\theta,n} = \frac{\theta e^{-\theta}}{\beta + 1 - e^{-\theta}} = \frac{\theta}{(\beta+1)e^{\theta} - 1} \tag{29}$$

where we have again used Fact 1. This limit for a large number of users is useful because now the total throughput is in terms of only two parameters β and θ, rather than three parameters β, θ, n. Given $\beta = 1/m \in (0,1)$, the value of $\theta > 0$ can be optimized numerically (it is difficult to obtain a closed form result).

It has been noticed that if β is much smaller than 1 ($\beta << 1$), the expression on the right-hand-side of (29) is "approximately" maximized by the following (see [2]):

$$\theta^* \approx \sqrt{2\beta} \quad (\beta << 1)$$
$$\lambda^* \approx \frac{1}{1+\sqrt{2\beta}} \quad (\beta << 1)$$

Notice that if $\beta \to 0$ then $\lambda^* \to 1$. Of course, the throughput can also be pushed to 1 via basic Aloha with capture, as described in the previous subsection. However, the use of mini-slots can significantly reduce the delay.

To understand the approximate analysis, first suppose θ is very small so that a truncated Taylor expansion gives $e^{\theta} \approx 1 + \theta + \theta^2/2$. The right-hand-side in (29) becomes

$$\underbrace{\frac{\theta}{(\beta+1)e^{\theta}-1}}_{f(\theta)} \approx \underbrace{\frac{\theta}{(\beta+1)(1+\theta+\theta^2/2)-1}}_{\tilde{f}(\theta)}$$

where $f(\theta)$ is denotes the exact function and $\tilde{f}(\theta)$ is the approximating function. The maximizer of $\tilde{f}(\theta)$ over all $\theta > 0$ is found at the point of zero derivative:

$$\frac{d}{d\theta}\tilde{f}(\theta) = 0 \implies [(\beta+1)(1+\theta+\theta^2/2)-1] = \theta(\beta+1)(1+\theta)$$

This is a quadratic equation in θ and the maximizer of $\tilde{f}(\theta)$ is

$$\tilde{\theta}^* = \frac{\sqrt{2\beta}}{\sqrt{1+\beta}}$$
$$\tilde{\lambda}^* = \tilde{f}(\tilde{\theta}^*) = \frac{1 - \frac{\sqrt{2\beta}}{\sqrt{1+\beta}}}{1-\beta}$$

These values $\tilde{\theta}^*$ and $\tilde{\lambda}^*$ give a better approximation than $\sqrt{2\beta}$ and $1/(1+\sqrt{2\beta})$. However, $\sqrt{2\beta}$ and $1/(1+\sqrt{2\beta})$ are simpler expressions that have been historically used for CSMA rules of thumb. Note that when $\beta \approx 0$ we have $\sqrt{2\beta}/\sqrt{1+\beta} \approx \sqrt{2\beta}$. The following table compares the approximations for various β values (Fig. 8).

β	Numerical search for exact (θ^*, λ^*)	$\left(\frac{\sqrt{2\beta}}{\sqrt{1+\beta}}, \frac{1-\frac{\sqrt{2\beta}}{\sqrt{1+\beta}}}{1-\beta}\right)$	$\left(\sqrt{2\beta}, \frac{1}{1+\sqrt{2\beta}}\right)$
0.1	(0.3755, 0.6245)	(0.4264, 0.6373)	(0.4472, 0.6910)
0.05	(0.2807, 0.7193)	(0.3086, 0.7278)	(0.3162, 0.7597)
0.025	(0.2061, 0.7939)	(0.2209, 0.7991)	(0.2236, 0.8173)
0.01	(0.1345, 0.8655)	(0.1407, 0.8680)	(0.1414, 0.8761)
0.001	(0.0440, 0.9560)	(0.0447, 0.9563)	(0.0447, 0.9572)

Fig. 8. Comparison of exact and approximate (θ^*, λ^*) values for slotted CSMA for various mini-slot sizes β.

I. Collision Detection (CSMA/CD)

Now suppose that users can sense the carrier for early detection of idle slots, but they can also sense for early detection of collision slots. Then collision slots can also be reduced. This gives rise to slotted CSMA/CD: The worst-case time to detect idle and collision slots is assumed to be $\beta = 1/m \in (0,1)$ and time is again slotted with respect to mini-slots of size β. The n users again independently transmit with probability p at the start of every renewal frame, but the new renewal structure is:

- Success: Renewal size is $\beta + 1$ time units, 1 packet is delivered. $(T_i, G_i) = (\beta+1, 1)$.
- Collision: Renewal size is 2β time units, 0 packets are delivered. $(T_i, G_i) = (2\beta, 0)$.
- Idle: Renewal size is β time units, 0 packets are delivered. $(T_i, G_i) = (\beta, 0)$.

The values of $P[Success]$, $P[Collision]$, $P[Idle]$ are the same as in CSMA. By renewal-reward theory the throughput is:

$$\lambda = \frac{\mathbb{E}[G_1]}{\mathbb{E}[T_1]} = \frac{P[Success]}{(1+\beta)P[Succ] + (2\beta)P[Collision] + \beta P[Idle]}$$

Similarly, we can let $p = \theta/n$ and let $n \to \infty$ to obtain

$$\lambda = \underbrace{\frac{\theta e^{-\theta}}{2\beta + \theta e^{-\theta} - \beta e^{-\theta} - \beta\theta e^{-\theta}}}_{f(\theta)} \quad (30)$$

Intuitively, we expect that p^* for CSMA/CD will be larger than the corresponding optimal transmit probability for CSMA because collision detection means that collisions incur less cost: We should transmit more aggressively. The reader can verify this is true via a numerical optimization.

Directly differentiating $f(\theta)$ as defined by the right-hand-side of the above inequality gives

$$\frac{d}{d\theta} f(\theta) = \frac{\beta(2e^{\theta}(\theta - 1) + 1)}{(\theta - \beta(\theta + 1 - 2e^{\theta}))^2}$$

which, remarkably, attains the value 0 when

$$2e^{\theta}(\theta - 1) + 1 = 0$$

which is independent of the value of β. A numerical search for the root gives

$$\theta^* = 0.7680$$

which is the maximizer of $f(\theta)$, and so

$$\lambda^* = f(\theta^*) \approx \frac{1}{1 + 3.31\beta}$$

Since $\theta = np$ is the expected number of transmitters at the start of a frame, at optimality we want the expected number of transmitters to be $\theta^* = 0.7680$.

Comparing with the CSMA throughput from the previous section (which holds for $\beta << 1$):

$$\lambda_{CSMA} \approx \frac{1}{1 + \sqrt{2\beta}}$$
$$\lambda_{CSMA/CD} \approx \frac{1}{1 + 3.31\beta}$$

when $\beta << 1$ then $\sqrt{2\beta} >> 3.31\beta$ and so the optimized throughput of CSMA/CD is indeed larger than that of CSMA.

J. *Unslotted Aloha*

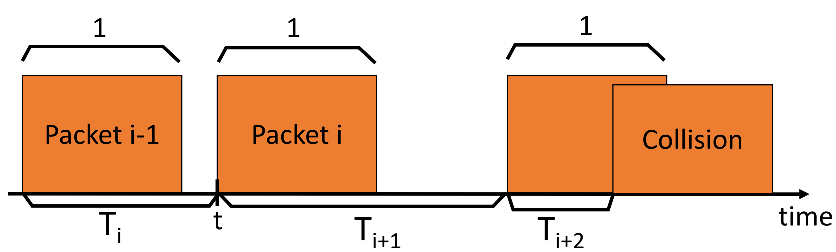

Fig. 9. A timing diagram for unslotted Aloha. Packet i arrives at time t and does not collide with either packet $i-1$ or packet $i+1$ (hence, $G_i = 1$). However, packet $i+1$ collides with packet $i+2$ and so $G_{i+1} = G_{i+2} = 0$.

In many systems it is difficult to implement a time slot structure. Consider a system with n users, each user has infinitely many packets to send, all packets take one unit of time, but there are no time slots. A user who transmits at some real-valued time $t \geq 0$ will overlap with another transmission in two possible ways:

- Overlap from the left: The packet collides with a previously transmitted packet. This takes place exactly when the previous packet initiated its transmission less than 1 unit of time before.
- Overlap from the right: The packet collides with a future packet. This takes place exactly when the next packet is transmitted less than one unit of time after.

For simplicity of analysis, we define a packet to *collide* if its transmission overlaps with *any part* of another transmission. For simplicity it is assumed that no data is received if there is a collision. The fact that collisions can take place *from both the right and from the left* will reduce throughput by a factor of two in comparison with slotted Aloha.

To define unslotted Aloha, start the system at time $t = 0$. Suppose each user $i \in \{1, \dots, n\}$ transmits according to an independent Poisson process of rate μ/n, where $\mu > 0$ is a parameter that shall be sized later to maximize throughput.[9] It is known that n independent Poisson processes form a single Poisson process with the sum rate. Thus, the collective transmission process is Poisson of rate μ. Let $N(t)$ count the total arrivals up to time t, and let $\{T_k\}_{k=1}^{\infty}$ denote the i.i.d. $exp(\mu)$ inter-transmission times associated with $N(t)$. Define the reward $R(t)$ as the total number of successes:

$$R(t) = \sum_{i=1}^{N(t)} G_i$$

where $G_i = 1$ if and only if transmission i is a success (see Fig. 9):

$$G_i = 1_{\{T_i > 1\} \cap \{T_{i+1} > 1\}} \quad \forall i \geq 2 \tag{31}$$

where 1_A is a binary-valued indicator function that is 1 if and only if event A is true, and

$$P[G_i = 1] = P[T_i > 1]P[T_{i+1} > 1] = e^{-\mu}e^{-\mu} = e^{-2\mu} \quad \forall i \geq 2$$

where we restrict attention to $i \geq 2$ because the first transmission cannot experience a collision from the left. The first transmission does not impact the long term throughput and so one can guess:

$$\lim_{t \to \infty} \frac{R(t)}{t} = \mu P[G_2 = 1] = \mu e^{-2\mu} \quad \text{(w.p.1)}$$

which is maximized over all $\mu > 0$ at $\mu^* = 1/2$ to give

$$\lambda^* = \frac{1}{2e}$$

which is indeed exactly a factor of 2 less than the slotted Aloha throughput of $1/e$.

How can this be rigorously derived? The main problem is that the random sequence $\{G_i\}_{i=1}^{\infty}$ defined in (31) is *not i.i.d.* and so renewal-reward theory cannot be directly used. However, we observe that the odd rewards $\{G_{2i-1}\}_{i=2}^{\infty}$ are i.i.d. amongst themselves (excluding G_1), and the even rewards $\{G_{2i}\}_{i=1}^{\infty}$ are i.i.d. amongst themselves. Define $M(t)$ as the number of complete *pairs* of rewards that occur up to time t and observe that

$$\sum_{i=1}^{M(t)} (G_{2i-1} + G_{2i}) \leq \sum_{j=1}^{N(t)} G_j \leq \sum_{i=1}^{M(t)+1} (G_{2i-1} + G_{2i})$$

Fix $t \geq T_1 + T_2$ so that $M(t) \geq 1$. Dividing both sides by t gives

$$\frac{1}{t}\sum_{i=1}^{M(t)} G_{2i-1} + \frac{1}{t}\sum_{i=1}^{M(t)} G_{2i} \leq \frac{R(t)}{t} \leq \frac{1}{t}\sum_{i=1}^{M(t)+1} G_{2i-1} + \frac{1}{t}\sum_{i=1}^{M(t)+1} G_{2i} \tag{32}$$

Since $M(t)$ have i.i.d. frame sizes of expected size $2\mathbb{E}[T_i] = 2/\mu$, and since the even rewards $\{G_{2i}\}_{i=1}^{\infty}$ are i.i.d. Bernoulli random variables with expectation $e^{-2\mu}$, we have by standard renewal-reward theory

$$\lim_{t \to \infty} \frac{1}{t}\sum_{i=1}^{M(t)} G_{2i} = \lim_{t \to \infty} \frac{1}{t}\sum_{i=1}^{M(t)+1} G_{2i} = \frac{\mathbb{E}[G_2]}{2/\mu} = \frac{1}{2}\mu e^{-2\mu} \quad \text{(w.p.1)}$$

Similarly, since G_1 has its own distribution but $\{G_{2i-1}\}_{i=2}^{\infty}$ are i.i.d. Bernoulli with expectation $e^{-2\mu}$, we have by the delayed renewal-reward theorem

$$\lim_{t \to \infty} \frac{1}{t}\sum_{i=1}^{M(t)} G_{2i-1} = \lim_{t \to \infty} \frac{1}{t}\sum_{i=1}^{M(t)+1} G_{2i-1} = \frac{\mathbb{E}[G_3]}{2/\mu} = \frac{1}{2}\mu e^{-2\mu} \quad \text{(w.p.1)}$$

Using these in (32) confirms our guess:

$$\lim_{t \to \infty} \frac{R(t)}{t} = \mu e^{-2\mu} \quad \text{(w.p.1.)}$$

[9]Strictly speaking, a single user who transmits according to a Poisson process of rate μ/n may send a packet less than one unit of time from its *own previous transmission* and hence could collide with itself. This happens a negligible fraction of time when n is large and so, for simplicity, we do not modify the transmissions for these cases.

K. Delayed feedback and parallel systems

Consider the slotted Aloha, k-slot capture, and ZigZag protocols, all of which use a fixed slot system (with no mini-slots). Suppose the idle/success/collision feedback is sent directly from the receiver over a control channel. Suppose the feedback is delayed, but the delay is at most m slots. In particular, suppose that feedback on transmissions of each slot $t \in \{0, 1, 2, \ldots\}$ is received before the start of slot $t+m$. An easy way to exactly implement the protocol is to separate the system into m parallel systems. System $i \in \{0, \ldots, m-1\}$ runs an exact version of the protocol over time slots

$$t \in \{i, i+m, i+2m, i+3m, \ldots\}$$

Thus, when $i \neq j$, the different systems i and j use disjoint time slots. The m different systems do not interfere with each other. The throughput of each system $i \in \{0, \ldots, m-1\}$ is equal to the throughput it would have without any feedback delay, *divided by* m (because throughput is now measured in terms of successes per m slots, rather than successes per slot). Total throughput (summed over all m systems) is exactly the same as it would be if there were no feedback delay. As a practical detail, the value of m could be included in the feedback so that newly arriving users are informed about the value of m before their first transmission.

L. Dynamically changing the transmit probability

The slotted Aloha idea of having everyone transmit with probability p is nice because it requires less coordination than round-robin. Unfortunately there are still several problems:

1) How do we get everyone to agree to use the same probability p?
2) How can we use the optimal probability $p^* = 1/n$ when the number of users n is unknown?
3) What happens if the number of users n changes over time?

One way to resolve these issues is to have each network user form an *estimate* of either n or $1/n$ based on a window of observed history.[10] One heuristic method to estimate n is to use the formula

$$P[Idle] = (1-p)^n \tag{33}$$

which holds under the special case assumption that there are n users that independently transmit with probability p. Now, that special case does not hold in the actual system. Thus, equation (33) is not intended to hold in the actual system. *Rather, equation* (33) *is used only to provide intuition*: Suppose all users keep a running time average idle probability $\overline{I}[t]$, which is the average number of idle slots within some window of past history (say, $\overline{I}[t]$ is the total number of idle slots in the last 10 transmissions, divided by 10, with a reasonable modified definition for slots $t < 10$). Each user $i \in \{1, \ldots, n\}$ knows its own average transmission probability $\overline{p}_i[t]$ (again averaged over the same window of history). Unfortunately, the $\overline{p}_i[t]$ values may be different across different users $i \in \{1, \ldots, n\}$. Nevertheless if each user i approximates all other users as having the same averaged probability, then user i can substitute its values for $\overline{I}[t]$ and $\overline{p}_i[t]$ as replacements for $P[Idle]$ and p in (33) to obtain the following approximation:

$$\overline{I}[t] \approx (1-\overline{p}_i[t])^n \implies n[t] \approx \frac{\log(\overline{I}[t])}{1-\overline{p}_i[t]}$$

where $n[t]$ is intended to be a (heuristic) approximation for the number of users in the network on slot t. This approximation has singularities when $\overline{I}[t] = 0$ and when $\overline{p}_i[t] = 1$ and so the approximation must be modified in those cases.

This is a simple and crude approximation. Consider an alternative max-likelihood approach: Suppose we observe the feedback $F[t] \in \{0, 1, C\}$ over the past 5 slots $\tau \in \{t, t-1, ..., t-4\}$ and we observe the following particular event A:

$$A = \{(F[t-4], F[t-3], F[t-2], F[t-1], F[t]) = (0, 0, 1, 0, C)\}$$

If we assume that on each slot t, all users i somehow use the same probabilities, so that $p_i[t] = p[t]$ for all $i \in \{1, \ldots, n\}$ (for some value $p[t]$), then the probability $P[A|n]$ for this event (assuming there were n users on each of the slots) is:

$$P[A|n] = (1-p[t-4])^n(1-p[t-3])^n(np[t-2](1-p[t-2])^{n-1})(1-p[t-1])^n(1-(1-p[t])^n - np[t](1-p[t])^{n-1})$$

and the max likelihood estimator for n is:

$$\hat{n}[t] = \arg\max_{n \in \{0,1,2,\ldots\}} P[A|n]$$

This is a more complex estimator but is (hopefully) more accurate. Of course, if users can have different transmit probabilities on the same slot t, then the assumption $p[t] = p_i[t]$ for all $i \in \{1, \ldots, n\}$ is itself an approximation. Given that we are already approximating, it may be reasonable to use a simpler approximation.

Another approach is for the users to implement some form of probability update (without attempting to estimate n or $1/n$). Each user i in the network on slot t can define a value $p_i[t] \in [0, 1]$ as the transmit probability it will use on slot t. After every

[10] Recall various types of estimators, such as MMSE, linear MMSE, Max a-priori, and Max-likelihood. Also recall that estimating a parameter n is not always the same as estimating $1/n$.

slot t, each user i can observe the window of history (including the 0/1/C feedback for slot t) and define $p_i[t+1]$ as either higher, lower, or the same as $p_i[t]$. Different update rules may lead to each different user having a different value of $p_i[t]$. Most rules that are implemented are *well reasoned heuristics* that perform well in experiments (see also [3] for a mathematical analysis that shows $1/e$ throughput for an estimation algorithm in a related slotted Aloha model).

For example, suppose all users agree in advance to use probabilities in the interval $[p_{min}, p_{max}]$ for some values $0 \leq p_{min} < p_{max} \leq 1$. Then each user i may want to update $p_i[t+1]$ based on the 0/1/C feedback from slot t as follows:

- If slot t is a collision then *decrease* the transmit probability:
$$p_i[t+1] = \max[p_i[t] - d_i[t], 0]$$
where $d_i[t]$ is some nonnegative value that possibly depends on $p_i[t]$ and/or on other historical information.
- If slot t is an idle or success then *increase* the probability:
$$p_i[t+1] = \min[p_i[t] + a_i[t], 1]$$
where $a_i[t]$ is some nonnegative value that possibly depends on $p_i[t]$ and/or on other historical information.

Other rules might distinguish the "idle" and "success" cases, and still others may use more detailed information such as running averages of collision and success history, rather than just the 0/1/C feedback from slot t.

Usually, one hopes that the rule for updating $p_i[t]$ has the property that if the number of users n is an unknown but fixed quantity that never changes, then
$$\lim_{t\to\infty} p_i[t] = 1/n \quad \forall i \in \{1, \ldots, n\}$$
Furthermore, if the number of users varies between $10 - 12$ for the first minute, and then abruptly changes to vary between $20 - 25$ in the next minute, one hopes the algorithm will allow $p_i[t]$ to quickly converge to $\approx 1/10$ for the first minute, and then change to $\approx 1/20$ in the next minute.

This sets the stage for a number of interesting computer experiments than can be conducted. Of course, when performing a computer simulation, it is still useful to precisely define the *mathematical model* that is being simulated. This is explored in the next section.

VI. Mathematical models for dynamic user behavior

Suppose we want to simulate a scheme for users to dynamically adjust their transmit probabilities $p_i[t]$, as discussed in Section V-L. The basic setting that can be considered first is when the number of users n is fixed, never changes over time, but is unknown to the users at time $t = 0$. This is an important scenario and can provide much intuition and insight into different methods for updating the $p_i[t]$ probabilities. However, eventually one may want to consider the important scenario where the number of users change. How should the changes be modeled? This section specifies three precise mathematical models.

A. Markov-based activity pattern

Suppose there are a fixed number of users n that never change. However, the users do not always want to send data. Suppose that each user $i \in \{1, \ldots, n\}$ has an *activity state* $A_i[t] \in \{0, 1\}$. If $A_i[t] = 1$ then the user is *active* and wants to send data, while $A_i[t] = 0$ means the user is idle (or "asleep") and does not want to send data. For each user $i \in \{1, \ldots, n\}$, the value value $A_i[t]$ typically stays 1 for long bursts of time t, and then switches to 0 for another long burst of time t. This models a situation where the number of users stays fixed for a long period of time, but those users do not always want to send data.

No system lasts for an infinite amount of time. However, mathematical models that assume an infinite number of time slots are often reasonable when a system is analyzed over a finite time that includes thousands of time slots. For example, imagine a coffee shop where, for a period of 20 minutes, there are exactly n people sitting at tables. Because a timeslot for packet transmission an be very small, there can be thousands of time slots during this 20 minute interval. The use of a binary activity state $A_i[t]$ for each user allows treatment of the real situation that the n people are not always using their wireless devices: Sometimes they are talking (face to face) with each other; sometimes they are simply drinking coffee! Thus, the total number of users that want to send data at a given time t can be defined as the *random process* $N[t] = \sum_{i=1}^{n} A_i[t]$. If each active user i knows the exact value of $N[t]$ at the start of each new slot t, then it could set its transmit probability to $p_i^*[t] = 1/N_i[t]$, which corresponds to the optimal common random transmit probability for slotted Aloha (as described in the previous section). On the other hand, in general the users may not know $N[t]$ and this scenario is useful for testing the robustness of various dynamic probability schemes (such as those described in Section V-L).

To conduct a computer simulation of this situation, we need a mathematical model to describe how $A_i[t]$ evolves for each user $i \in \{1, \ldots, n\}$. One simple but useful model is to assume that the activity state of each user evolves independently according to a 2-state discrete time Markov chain (DTMC) with transition probabilities ϵ_i and δ_i, as shown in Fig. 3. It is assumed that $0 < \epsilon_i < 1$ and $0 < \delta_i < 1$. At the end of each slot t the activity state of each user $i \in \{1, \ldots, n\}$ is updated independently as follows:

- If $A_i[t] = 0$ then independently flip a biased coin with $P[HEADS] = \epsilon_i$ and update:

$$A_i[t+1] = \begin{cases} 1 & \text{if HEADS} \\ 0 & \text{otherwise} \end{cases}$$

- If $A_i[t] = 1$ then independently flip a biased coin with $P[HEADS] = \delta_i$ and update:

$$A_i[t+1] = \begin{cases} 0 & \text{if HEADS} \\ 1 & \text{otherwise} \end{cases}$$

It is important to understand the mathematical properties of this activity process: The timeline for each user i can be decomposed into IDLE slots and ACTIVE slots. The duration of each IDLE period is independent of history and is geometrically distributed with parameter ϵ_i. Thus, the average duration of a user i IDLE slot is $1/\epsilon_i$. Similarly the duration of each user i ACTIVE period i s independent of history and is geometrically distributed with parameter δ_i, having average size $1/\delta_i$. Using the cut-set equations it is easy to determine the *steady state probabilities* $\pi_i(ACTIVE)$ and $\pi_i(IDLE)$ for each user i:

$$\pi_i(ACTIVE) = \lim_{t\to\infty} P[A_i[t] = 1] = \frac{\epsilon_i}{\epsilon_i + \delta_i}$$
$$\pi_i(IDLE) = \lim_{t\to\infty} P[A_i[t] = 0] = \frac{\delta_i}{\epsilon_i + \delta_i}$$

and these limits are the same regardless of the initial state $A_i[0] \in \{0, 1\}$. These steady state probabilities can also be shown to satisfy

$$\pi_i(ACTIVE) = \frac{\mathbb{E}\left[ACTIVE_i\right]}{\mathbb{E}\left[ACTIVE_i + IDLE_i\right]}$$
$$\pi_i(IDLE) = \frac{\mathbb{E}\left[IDLE_i\right]}{\mathbb{E}\left[ACTIVE_i + IDLE_i\right]}$$

where $\mathbb{E}\left[ACTIVE_i\right] = 1/\delta_i$ and $\mathbb{E}\left[IDLE_i\right] = 1/\epsilon_i$. These equations will be illuminated in terms of *renewal theory* as discussed in the next section.

Under this model, the average number of users that are active on a given slot t is

$$\mathbb{E}\left[N[t]\right] = \sum_{i=1}^{n} \mathbb{E}\left[A_i[t]\right] = \sum_{i=1}^{n} P[A_i[t] = 1]$$

which uses the fact

$$A_i[t] \in \{0, 1\} \implies \mathbb{E}\left[A_i[t]\right] = 1 \cdot P[A_i[t] = 1] + 0 \cdot P[A_i[t] = 0] = P[A_i[t] = 1]$$

The steady state average number of active users is thus:

$$\lim_{t\to\infty} \mathbb{E}\left[N[t]\right] = \sum_{i=1}^{n} \lim_{t\to\infty} P[A_i[t] = 1]$$
$$= \sum_{i=1}^{n} \frac{\epsilon_i}{\epsilon_i + \delta_i}$$

A simple simulation setting is when we use the same ϵ_i and δ_i parameters for all users i, so that $\epsilon_i = \epsilon$ and $\delta_i = \delta$ for some values $\epsilon \in (0, 1)$ and $\delta \in (0, 1)$, where δ is chosen to obtain a desired average active period of $1/\delta$ slots, ϵ is chosen to obtain a desired average idle period of $1/\delta$ slots, and the average number of active users is $n\epsilon/(\epsilon + \delta)$.

Consider a computer simulation of this model that uses some particular multi-access method at each user. Let $X_i[t] \in \{0, 1\}$ denote the user i success process on slot t, as defined in (22). Suppose the simulation runs for $T = 100 \times 10^3$ slots. The computer simulation allows us to compute the throughput vector $(\overline{X}_1[T], \ldots, \overline{X}_n[T])$ and the corresponding total throughput $\overline{X}[T]$ and the Jain fairness index $F(\overline{X}_1[T], \ldots, \overline{X}_n[T])$. If all users have a symmetric activity behavior, so that $\epsilon_i = \epsilon$ and $\delta_i = \delta$ for all $i \in \{1, \ldots, n\}$, then each user is active for the same fraction of time and it is a reasonable desire the vector $(\overline{X}_1, \ldots, \overline{X}_n)$ to have near-equal components (as indicated by a Jain fairness index that is close to 1). However, if the activity behavior of the users is *not* symmetric then there is a major problem with the Jain fairness index itself: It is not fair! Specifically, suppose there are three users with different steady state ACTIVE probabilities as follows:

$$\pi_1(ACTIVE) = 0.9, \pi_2(ACTIVE) = 0.5, \pi_3(ACTIVE) = 0.1$$

Since user 3 is only active 10 percent of the time, *the largest throughput that can be achieved by user 3 is* 0.1. Therefore, the most objectively fair throughput vector to strive for in this situation is:

$$(\overline{X}_1, \overline{X}_2, \overline{X}_3) = (0.45, 0.45, 0.1)$$

Indeed, this gives the best possible total throughput $\overline{X} = \overline{X}_1 + \overline{X}_2 + \overline{X}_3 = 1$; gives user 3 its largest possible throughput $\overline{X}_3 = 0.1$; gives users 1 and 2 an equal share of the remaining throughput. However, the Jain fairness index yields:

$$F(0.45, 0.45, 0.1) = \frac{(0.45 + 0.45 + 0.1)^2}{3(0.45^2 + 0.45^2 + 0.1^2)} \approx 0.803$$

The fact that the Jain fairness index is only 0.803 (not 1) *incorrectly suggests* that fairness can be improved.

B. Randomly arriving users that stay for a finite time

An alternative model is to assume that there are an infinite number of users that arrive over time according to a *random process*. Each user arrives at a certain time slot, stays for a random amount of time, and then leaves (never to return). It is assumed that the each user has an infinite amount of data that it wants to send while it is in the system (so that it always has a packet to send when it is in the system). Let $A[t]$ be the number of new arrivals on slot $t \in \{0, 1, 2, \ldots\}$. For simplicity assume that $\{A[t]\}_{t=0}^{\infty}$ is i.i.d. over slots with some mass function $P[A[t] = i]$ for $i \in \{0, 1, 2, \ldots\}$ that satisfies

$$\sum_{i=0}^{\infty} P[A[t] = i] = 1$$

Define λ as the average number of new arrivals per slot:

$$\lambda = \mathbb{E}\left[A[t]\right] = \sum_{i=0}^{\infty} iP[A[t] = i]$$

By the *law of large numbers* we know

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} A[t] = \lambda \quad \text{(with prob 1)}$$

For example, one might use a *Poisson mass function* with parameter $\lambda > 0$ so that:

$$P[A[t] = i] = \frac{\lambda^i}{i!}e^{-\lambda} \quad \forall i \in \{0, 1, 2, \ldots\}$$

Alternatively, one might use a *Bernoulli mass function* so that there can be at most one new arrival per slot:

$$P[A[t] = 1] = \lambda, \quad P[A[t] = 0] = 1 - \lambda$$

Assume that each arriving user i stays in the system for an independent amount of time slots W_i, where $\{W_i\}_{i=1}^{\infty}$ is an i.i.d. sequence of positive integers that is independent of the arrival sequence $\{A[t]\}_{t=0}^{\infty}$. For simplicity of notation, let W be an independent copy of W_1. The simplest model is when W is a geometric random variable with parameter q, so that

$$P[W = k] = q(1-q)^{k-1} \quad \forall k \in \{1, 2, 3, \ldots\}$$

in which case

$$\mathbb{E}\left[W\right] = \sum_{k=0}^{\infty} kq(1-q)^{k-1} = 1/q$$

This geometric staying time model is the easiest to simulate because of its *memoryless property*: At the end of every slot t, each user that is currently in the system independently decides to leave with probability q (and to stay with probability $1-q$).

Let $N[t]$ be the number of users currently in the system on slot t. Define $\overline{N}$ as the time average:

$$\overline{N} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} N[t]$$

Can we mathematically compute $\overline{N}$? This is a deep question and the answer is not obvious. It turns out that the answer is "yes." The answer does not require knowledge of the distribution of $A[t]$ or the distribution of W, it only requires their *average* values $\mathbb{E}\left[A[t]\right] = \lambda$ and $\mathbb{E}\left[W\right] = 1/q$. Indeed, a general *queueing theory* result called *Little's theorem* (which shall be studied in more detail later in the course) implies that, with probability 1,

$$\overline{N} = \lambda\mathbb{E}\left[W\right] = \lambda/q$$

C. Queue-based models

Again suppose there are a fixed number of users n that never changes over time. However, instead of modeling the active/busy behavior of the users according to 2-state DTMCs as in Section VI-A, suppose each user $i \in \{1, \ldots, n\}$ has data that it wants to send arrive according to an independent *arrival process*. The arriving data for user i is stored in a queue at user i. The *effective* number of users in the system is equal to the number of users that have non-empty queues.

Specifically, let $A_i[t]$ be the number of new packets that arrive to user i on slot t. Assume that the processes $\{A_i[t]\}_{t=0}^{\infty}$ are independent across users $i \in \{1, \ldots, n\}$ and are i.i.d. over slots with some probability mass function $P[A_i[t] = k]$ that may be different for each user i and that satisfies the following for all slots t and all users i:

$$\sum_{k=0}^{\infty} P[A_i[t] = k] = 1$$

Define λ_i as the expected number of new arrivals for user i on a given slot:

$$\lambda_i = \mathbb{E}\left[A_i[t]\right] = \sum_{k=0}^{\infty} kP[A_i[t] = k]$$

By the law of large numbers we have

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} A_i[t] = \lambda_i \quad \text{(with prob 1)} \tag{34}$$

Define $X_i[t] \in \{0, 1\}$ as in (22), so that $X_i[t] = 1$ if and only if user i successfully transmits a packet on slot t. Let $Q_i[t]$ be a nonnegative integer that represents the *queue backlog* for user i on slot t. Assume the queue is initially empty, so that $Q_i[0] = 0$. The queue update equation is:[11]

$$Q_i[t+1] = \max[Q_i[t] - X_i[t], 0] + A_i[t] \quad \forall t \in \{0, 1, 2, \ldots\} \tag{35}$$

Since $Q_i[0]$, $X_i[t]$ and $A_i[t]$ are always nonnegative integers, the value of $Q_i[t]$ is a nonnegative integer for all $t \in \{0, 1, 2, \ldots\}$. If $Q_i[t] > 0$ on slot $t \in \{0, 1, 2, \ldots\}$ then user i can attempt to transmit a packet (using some transmission probability $p_i[t]$). One can view the new packets arrivals $A_i[t]$ as arriving either in the middle or at the end of a slot t, so that they cannot be used for transmission on slot t. However, these new arrivals on slot t are added to the backlog and are available for transmission on slot $t + 1$. Let $N[t]$ be the number of non-empty queues at the start of slot t:

$$N[t] = \sum_{i=1}^{n} 1_{\{Q_i[t]>0\}}$$

where $1_{\{Q_i[t]>0\}}$ is called an *indicator function* that is 1 if $Q_i[t] > 0$ and 0 else:

$$1_{\{Q_i[t]>0\}} = \begin{cases} 1 & \text{if } Q_i[t] > 0 \\ 0 & \text{otherwise} \end{cases}$$

The value $N[t]$ is the number of users in the system on slot t that compete for access. Define $\overline{N}$ as its time average:

$$\overline{N} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} N[t]$$

Unlike the previous two models, the value of $\overline{N}$ now depends on the success history of the multi-access scheme. It is not obvious how to compute $\overline{N}$.

This queueing model adds a new mathematical topic of *queue stability* to the problem: The multi-access scheme needs to be good enough to provide each user i enough successes so that the packets that go *into* queue i eventually also get served. If there are not enough successes then the queue backlog can go to infinity! To see this, fix $i \in \{1, \ldots, n\}$. From (35) we obtain

$$Q_i[t+1] \geq Q_i[t] - X_i[t] + A_i[t] \quad \forall t \in \{0, 1, 2, \ldots\}$$

Thus

$$Q_i[t+1] - Q_i[t] \geq A_i[t] - X_i[t] \quad \forall t \in \{0, 1, 2, \ldots\}$$

Fix T as a positive integer. Summing the above over $t \in \{0, 1, \ldots, T-1\}$ gives

$$Q_i[T] - \underbrace{Q_i[0]}_{0} \geq \sum_{t=0}^{T-1} A_i[t] - \sum_{t=0}^{T-1} X_i[t]$$

[11]If we assume user i cannot transmit anything on slot t when $Q_i[t] = 0$ then we necessarily have $X_i[t] = 0$ when $Q_i[t] = 0$ and the queue update equation (35) can be simplified to $Q_i[t+1] = Q_i[t] - X_i[t] + A_i[t]$. We leave the update equation in the form (35) because it is sometimes useful to consider $X_i[t]$ as a *virtual* service opportunity that can take place even when the queue backlog is zero.

Dividing both sides by T and using $Q_i[0] = 0$ gives

$$\frac{Q_i[T]}{T} \geq \frac{1}{T}\sum_{t=0}^{T-1} A_i[t] - \frac{1}{T}\sum_{t=0}^{T-1} X_i[t]$$

Taking a limit as $T \to \infty$ (assuming for simplicity that the limit exists) gives

$$\lim_{T\to\infty} \frac{Q_i[T]}{T} \geq \lambda_i - \underbrace{\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} X_i[t]}_{\overline{X}_i} \quad \text{(with prob 1)}$$

where we have used (34). Define $\overline{X}_i$ as the average success rate for user i, as defined in the underbrace of (36). Then

$$\lim_{T\to\infty} \frac{Q_i[T]}{T} \geq \lambda_i - \overline{X}_i \quad \text{(with prob 1)} \tag{36}$$

Since the $\lim_{T\to\infty} Q_i[T]/T \geq 0$, from (36) we must have

$$0 \leq \overline{X}_i \leq \lambda_i \quad \forall i \in \{1, \ldots, n\} \tag{37}$$

It follows from (36) that

$$\overline{X}_i < \lambda_i \implies \lim_{T\to\infty} Q_i[T] = \infty \quad \text{(with prob 1)}$$

The situation where the queue backlog grows to infinity with probability 1 is certainly an unstable situation! There are various useful definitions of queue stability in the literature [18]. One of the simplest is the definition of *rate stability* given below.

Definition 5: A queue processs $\{Q_i[t]\}_{t=0}^{\infty}$ of the form (35) is said to be *rate stable* if

$$\lim_{T\to\infty} \frac{Q_i[T]}{T} = 0 \quad \text{(with prob 1)}$$

The inequality (36) implies that the queue is rate stable if and only if $\overline{X}_i = \lambda_i$. In particular, if $\overline{X}_i < \lambda_i$ then queue i is *not* rate stable. To ensure that all queues are stable, it is imperative that the average success rate $\overline{X}_i$ be equal to the arrival rate λ_i for each user i. On the other hand, there is at most one successful transmission per slot and so

$$\sum_{i=1}^{n} X_i[t] \leq 1 \implies \sum_{i=1}^{n} \overline{X}_i \leq 1$$

Thus, if $\sum_{i=1}^{n} \lambda_i > 1$, it is *impossible* to have $\lambda_i = \overline{X}_i$ for all i: Regardless of the multi-access policy, there must be some queue that is unstable! This means that the sum demand for access exceeds the system capabilities: The only way to stabilize the user queues is to start dropping data, so that the queue update equation (35) is modified to include an additional *packet dropping* term.

Now consider a situation with three users and suppose $(\lambda_1, \lambda_2, \lambda_3) = (0.1, 0.2, 0.3)$. Suppose a multi-access strategy is used that stabilizes all queues, so that $(\overline{X}_1, \overline{X}_2, \overline{X}_3) = (\lambda_1, \lambda_2, \lambda_3)$. This is the best possible throughput vector because it gives each user a success rate equal to its data arrival rate. The total throughput is $\lambda_1 + \lambda_2 + \lambda_3 = 0.6$. Even though this is less than 1, it cannot be improved. Further, the Jain fairness index is:

$$F(\lambda_1, \lambda_2, \lambda_2) = \frac{(0.1 + 0.2 + 0.3)^2}{3(0.1^2 + 0.2^2 + 0.3^2)} \approx 0.857$$

Even though this value is less than 1, it cannot be improved.

VII. A Robbins-Monro approach to dynamically adjusting the transmit probability

Consider a simplification of the dynamic user models of the previous section: The number of users is a fixed but unknown parameter $n \in \{2, 3, \ldots, m\}$. It is assumed that $n \geq 2$ so that there are at least two users who compete for channel resources. The value m represents a known maximum value on the number of users. Users transmit packets to an access point over slotted time $t \in \{0, 1, 2, \ldots\}$. We assume the idle/success/collision model where a success occurs on a slot t if and only if exactly one user transmits on slot t.

If the number of users n were known, we could have all users independently transmit with probability $p^* = 1/n$ on each slot and the total throughput of $(1 - 1/n)^{n-1} \approx 1/e$ would be attained. Here we let $S[t]$ be an *estimate* of n that is made on slot t. For convenience, we allow $S[t]$ to be a noninteger value, but we assume $0 \leq S[t] \leq m$ on every slot t. Every slot t, the access point broadcasts the value $p[t] = 1/S[t]$ as the common probability that all users should use that slot. The goal is to design an estimation procedure that enables $S[t]$ to converge closely to n as $t \to \infty$. If convergence is fast enough, then, intuitively, our scheme should also do well when the number of users n varies with time according to some random process $N[t]$ (such as the processes described in the previous section), provided that $N[t]$ does not change too quickly and dramatically.

Dynamic procedures of this type are considered under related models in [19][20][21][3]. The method to be discussed in this section is most closely related to the Robbins-Monro learning technique [22][23][24][25][26][27].

A. Update structure

Consider the following update structure (similar to [20][21]): Fix constant $\epsilon > 0$ (to be chosen later). On slot $t = 0$, arbitrarily choose the initial estimate $S[0] \in [2, m]$ (for example, choose $S[0] = 2$). Define $S[t]$ for $t > 0$ according to the following update equation:

$$S[t+1] = [S[t] + \epsilon J[t]]_2^m \quad \forall t \in \{0, 1, 2, \ldots\} \tag{38}$$

where $[\cdot]_2^m$ denotes a projection onto the interval $[2, m]$, that is

$$[x]_2^m = \begin{cases} 2 & \text{if } x < 2 \\ x & \text{if } x \in [2, m] \\ m & \text{if } x > m \end{cases}$$

and where $J[t]$ is defined by

$$J[t] = \begin{cases} -a[t] & \text{if slot } t \text{ has an idle} \\ b[t] & \text{if slot } t \text{ has a success} \\ c[t] & \text{if slot } t \text{ has a collision} \end{cases} \tag{39}$$

where $a[t], b[t], c[t]$ are values that shall be chosen later. The value ϵ shall be called the *stepsize* and the function $J[t]$ shall be called the *adjustment*. Intuitively, if slot t has an idle then we assume the actual number of users is lower than our estimate and so we make $J[t] = -a[t]$ (typically a negative value so that our estimate will be reduced on slot $t+1$). On the other hand, if slot t has a collision then we assume the actual number of users is larger than our estimate and so we make $J[t] = c[t]$ (typically a positive value so that our estimate will be increased on slot $t+1$). Finally, if we have a success then it is not clear if we should increase or decrease our estimate, or keep it the same. Thus, we make $J[t] = b[t]$ in this case, where $b[t]$ is a (possibly zero) value that we shall have the freedom to choose later. It shall be assumed that $a[t], b[t], c[t]$ are either constants or are deterministic functions of the current estimate $S[t]$.

The goal is to size the values $a[t], b[t], c[t]$ so that

$$\mathbb{E}\left[J[t]|S[t]\right] \leq 0 \text{ if } S[t] \geq n \tag{40}$$

$$\mathbb{E}\left[J[t]|S[t]\right] \geq 0 \text{ if } S[t] < n \tag{41}$$

Intuitively, this means that the estimate $S[t]$ tends to *decrease* when it is currently larger than the true number of users n, and tends to *increase* otherwise. Of course, defining $a[t] = b[t] = c[t] = 0$ gives $\mathbb{E}\left[J[t]|S[t]\right] = 0$ for all slots t and all $S[t] \in [2, m]$, so that (40)-(41) are satisfied. However, this choice does not help because the feedback adjustment $J[t]$ is always zero and so $S[t] = S[0]$ for all t: there is no learning and no convergence to the true value n. Therefore, it is not enough that (40)-(41) are satisfied, we shall require these to be satisfied with *strict and sufficiently large inequality* when $S[t] \neq n$.

B. Computing $\mathbb{E}\left[J[t]|S[t]\right]$

Recall that on each slot t, all n users independently transmit with probability $p[t] = 1/S[t]$. The value $p[t]$ is known to all users because it is assumed the access point broadcasts this value to everyone. By definition of $J[t]$ in (39) we have:

$$\mathbb{E}\left[J[t]|S[t]\right] = -a[t](1-p[t])^n + b[t]np[t](1-p[t])^{n-1} + c[t](1-(1-p[t])^n - np[t](1-p[t])^{n-1})$$

where the above uses the assumption that $a[t], b[t], c[t]$ are deterministic functions of the state $S[t]$, and they act as known constants when conditioning on $S[t]$. Collecting terms gives

$$\mathbb{E}\left[J[t]|S[t]\right] = c[t] - (1-p[t])^n(a[t]+c[t]) + np[t](1-p[t])^{n-1}(b[t]-c[t])$$

C. A particular choice of $a[t], b[t], c[t]$

Observe that $np[t] - 1 = \frac{n}{S[t]} - 1 > 0$ if and only if $S[t] < n$. It makes sense to manipulate the above expression to get a factor $np[t] - 1$,

$$\mathbb{E}\left[J[t]|S[t]\right] = \left[c[t] - (1-p[t])^n \underbrace{\left(a[t] + c[t] + \frac{b[t]-c[t]}{1-p[t]}\right)}\right] + (np[t]-1)(1-p[t])^{n-1}\underbrace{(b[t]-c[t])}$$

The value $S[t]$ is known, the value $p[t] = 1/S[t]$ is known, but the value n is unknown. The goal is to choose $a[t], b[t], c[t]$ as deterministic functions of the known $S[t]$ so the terms in the underbraces above yield a simple and useful expression. It shall be useful to consider the following choices of $a[t], b[t], c[t]$:

$$a[t] = -1 + 2(1-p[t])^{-S[t]}$$

$$b[t] = 1 + (1-p[t])^{-S[t]+1}$$

$$c[t] = 1$$

These are indeed deterministic functions of the known state $S[t]$ (recall that $p[t] = 1/S[t]$). Note also that they are not functions of the unknown parameter n (the algorithm could not be implemented otherwise). It can be shown that whenever $2 \leq S[t] < \infty$, these choices ensure:

$$2e - 1 \leq a[t] \leq 7$$
$$3 \leq b[t] \leq 1 + e$$

Under these choices we have:[12]

$$\begin{aligned}\mathbb{E}\left[J[t]|S[t]\right] &= \left[1 - (1 - p[t])^{n-S[t]}\right] + (np[t] - 1)(1 - p[t])^{n-S[t]} \\ &= \left(1 - (1 - p[t])^{n-S[t]}\right) + p[t](n - S[t])(1 - p[t])^{n-S[t]} \end{aligned} \quad (42)$$

where the final equality uses $p[t] = 1/S[t]$. It can be shown that the right-hand-side of (42) satisfies (40)-(41) with strict inequality when $S[t] \neq n$. Indeed, observe that $0 < p[t] < 1$ for all t and consider the two cases:

- Suppose $S[t] > n$. Then $(1 - p[t])^{n-S[t]} > 1$ and $(n - S[t]) < 0$ and so
$$\left(1 - (1 - p[t])^{n-S[t]}\right) + p[t](n - S[t])(1 - p[t])^{n-S[t]} < 0$$
- Suppose $S[t] < n$. Then $(1 - p[t])^{n-S[t]} < 1$ and $(n - S[t]) > 0$ and so
$$\left(1 - (1 - p[t])^{n-S[t]}\right) + p[t](n - S[t])(1 - p[t])^{n-S[t]} > 0$$

D. The case $n = 1$

The algorithm $S[t + 1] = [S[t] + \epsilon J[t]]_2^m$ was developed under the assumption that $n \geq 2$. This algorithm restricts to using $S[t] \geq 2$ for all t, and hence $p[t] \leq 1/2$ on each slot t. A major shortcoming of this algorithm is that it does not adequately treat the case $n = 1$. Indeed, if $n = 1$ then, since this algorithm yields $p[t] \leq 1/2$ for all slots t, the throughput is at most $1/2$. However, if $n = 1$, a good algorithm should learn to nearly always transmit so that near-perfect throghput of 1 is achieved.

One approach is to simply change the interval of projection: Use the new interval $[1, m]$ instead of the old interval $[2, m]$. The problem with this approach is that if $S[t] = 1$ then $p[t] = 1/S[t] = 1$ and the $a[t]$ and $b[t]$ constants defined in the previous subsection have a singularity. In particular, observe that:

$$a[t] = -1 + \frac{2}{(1 - \frac{1}{S[t]})^{S[t]}}$$

and the denominator is 0 when $S[t] = 1$. Alternatively, one could take a projection onto the interval $[1 + \delta, m]$ for some value δ that satisfies $0 < \delta \leq 1$. In this case the $a[t]$ coefficient satisfies:

$$2e - 1 \leq a[t] = -1 + \frac{2}{(1 - \frac{1}{S[t]})^{S[t]}} \leq -1 + \frac{2}{(1 - \frac{1}{1+\delta})^{1+\delta}}$$

However, the right-hand-side bound can be very large when $\delta \approx 0$ because

$$\lim_{\delta \to 0^+} \left[-1 + \frac{2}{(1 - \frac{1}{1+\delta})^{1+\delta}}\right] = \infty$$

The reason that we want to keep the coefficients $a[t]$ and $b[t]$ bounded by small values is that a bound on $\mathbb{E}\left[J[t]^2\right]$ shall be important for analytical guarantees (see Theorem 4 in the next subsection).

Another approach is to primarily use the iteration $S[t + 1] = [S[t] + \epsilon J[t]]_2^m$, but if the current estimate $S[t]$ is close to 2 and we have not recently seen any collisions, we can *heuristically guess* that $n = 1$ and we can have an *experimental slot* that uses $p[t] = 1$ to test this guess. If this experimental slot results in a success, we *know* that $n = 1$ and we can continue using $S[t] = 1$ and $p[t] = 1$ indefinitely (until a collision occurs, where a collision indicates that the number of users has changed from $n = 1$ to $n > 1$). Otherwise, if this experimental slot results in a collision, we *know* that $n \geq 2$ and we continue to use the iteration $S[t + 1] = [S[t] + \epsilon J[t]]_2^m$ until a long enough time has elapsed (with $S[t]$ still close to 2) so that we want to try another experimental slot to see if the number of users has changed to 1.

[12]We are unaware of similar choices of $a[t], b[t], c[t]$ in the literature and we believe these are novel. For example, work in [20][21] uses fixed parameters a, b, c, work in [19] uses a multiplicative update rule, work in [3] uses a different update structure, and most prior work considers a different model with an infinite number of users, all with only one packet.

E. Learning-based estimation of a parameter

The estimation procedure (38) falls into a general class of *stochastic approximation algorithms*. In particular, it is similar to a classic Robbins-Monro approximation procedure [22] (see also [23][24][25][26][27]). In this subsection we present a theorem that allows such estimation procedures to be analyzed. We consider a general problem of estimating an unknown parameter s^* that takes values in some known interval $[s_{min}, s_{max}]$. For example, in the multi-access problem we can use $s_{min} = 2, s_{max} = m, s^* = n$. Let $\{S[t]\}_{t=0}^{\infty}$ be the sequence of estimates with initial condition $S[0] \in [s_{min}, s_{max}]$ and update equation:[13]

$$S[t+1] = [S[t] + \epsilon J[t]]_{s_{min}}^{s_{max}} \quad \forall t \in \{0, 1, 2, \ldots\} \tag{43}$$

where $[x]_{s_{min}}^{s_{max}}$ is a projection onto the interval $[s_{min}, s_{max}]$; $\epsilon > 0$ is a *stepsize parameter* to be chosen later; $J[t]$ is some type of adjustment function (compare (43) with (38) for the case $s_{min} = 2, s_{max} = m, s^* = n$).

The estimates $\{S[t]\}_{t=0}^{\infty}$ that arise from the update equation (43) can be analyzed when the conditional expectations of $J[t]$, given $S[t]$, have certain structural bounds. Specifically, let $g : \mathbb{R} \to \mathbb{R}$ be function that shall be called a *bounding function*. It is assumed that $g(x)$ has the following properties:

$$g(x) \geq 0 \text{ if } x < 0 \tag{44}$$

$$g(x) \leq 0 \text{ if } x > 0 \tag{45}$$

Suppose we can engineer an adjustment function $J[t]$ that has the following property:

$$\mathbb{E}\left[J[t]|S[t] = s\right] \leq g(s - s^*) \quad \text{ if } s \geq s^* \tag{46}$$

$$\mathbb{E}\left[J[t]|S[t] = s\right] \geq g(s - s^*) \quad \text{ if } s < s^* \tag{47}$$

A typical bounding function $g(x)$ is piecewise linear:

$$g(x) = \begin{cases} -\beta_1 x & \text{if } x \leq 0 \\ -\beta_2 x & \text{if } x > 0 \end{cases} \tag{48}$$

for some constants $\beta_1 > 0$ and $\beta_2 > 0$. Fig. 10 gives an example illustration of the inequalities (46)-(47) with respect to such a piecewise linear bounding function $g(x)$.

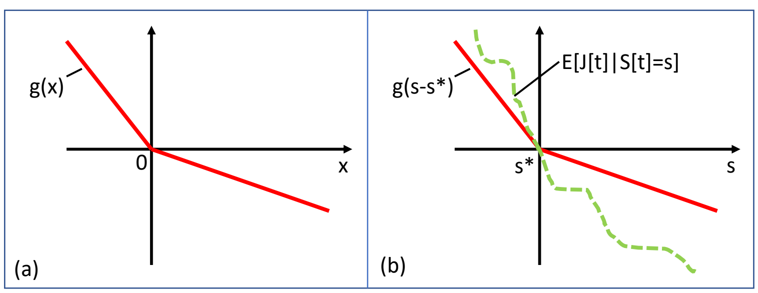

Fig. 10. (a) A plot of the piecewise linear bounding function $g(x)$ from (48); (b) An illustration of inequalities (46)-(47).

Theorem 4: (Learning-based estimation of s^*) Fix constants $\epsilon > 0$ and $S[0] \in [s_{min}, s_{max}]$. Suppose $S[t]$ evolves according to (43). Suppose there is a bounding function $g(x)$ that satisfies (44)-(45) such that $J[t]$ satisfies (46)-(47) for all slots $t \in \{0, 1, 2, \ldots\}$. Further suppose there is a real number $c \geq 0$ such that

$$\mathbb{E}\left[J[t]^2\right] \leq c \quad \forall t \in \{0, 1, 2, \ldots\} \tag{49}$$

Then

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left[|(S[t] - s^*)g(S[t] - s^*)|\right] \leq \frac{c\epsilon}{2} + \frac{(S[0] - s^*)^2}{2\epsilon T} \quad \forall T \in \{1, 2, 3, \ldots\} \tag{50}$$

In particular,

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left[|(S[t] - s^*)g(S[t] - s^*)|\right] \leq \frac{(c+1)\epsilon}{2} \quad \forall T \geq \frac{(b-a)^2}{\epsilon^2}$$

[13] To consider an unknown parameter s^* that can take values in the entire real number line $\mathbb{R}$, we use $s_{min} = -\infty$ and $s_{max} = \infty$ so the update (43) becomes $S[t+1] = S[t] + \epsilon J[t]$.

Proof: Fix T as a positive integer and fix $t \in \{0, 1, \ldots, T-1\}$. We have

$$\begin{aligned}
\frac{1}{2}(S[t+1]-s^*)^2 &\overset{(a)}{=} \frac{1}{2}([S[t]+\epsilon J[t]]_a^b - s^*)^2 \\
&\overset{(b)}{=} \frac{1}{2}([S[t]+\epsilon J[t]]_a^b - [s^*]_a^b)^2 \\
&\overset{(c)}{\leq} \frac{1}{2}(S[t]+\epsilon J[t] - s^*)^2 \\
&= \frac{1}{2}(S[t]-s^*)^2 + \frac{\epsilon^2 J[t]^2}{2} + \epsilon(S[t]-s^*)J[t]
\end{aligned}$$

where (a) holds by the update equation (43); (b) holds because $s^* \in [s_{min}, s_{max}]$ and so $s^* = [s^*]_a^b$; (c) holds by the well known convex analysis fact that the distance between the projections of two real numbers onto a closed interval $[s_{min}, s_{max}]$ is less than or equal to the original distance between these numbers. Taking expectations of the above gives

$$\frac{1}{2}\mathbb{E}\left[(S[t+1-s^*)^2\right] - \frac{1}{2}\mathbb{E}\left[(S[t]-s^*)^2\right] \leq \frac{\epsilon^2 c}{2} + \epsilon\mathbb{E}\left[(S[t]-s^*)J[t]\right] \tag{51}$$

where we have used (49). To compute $\mathbb{E}\left[(S[t]-s^*)J[t]\right]$ we condition on $S[t] = s$ for two cases:

- Suppose $s \geq s^*$. Then

$$\begin{aligned}
\mathbb{E}\left[(S[t]-s^*)J[t]|S[t]=s\right] &= (s-s^*)\mathbb{E}\left[J[t]|S[t]=s\right] \\
&\overset{(a)}{\leq} (s-s^*)g(s-s^*) \\
&\overset{(b)}{=} -|(s-s^*)g(s-s^*)|
\end{aligned}$$

  where (a) uses $(s-s^*) \geq 0$ and (46); (b) uses $(s-s^*) \geq 0$ and (45).
- Suppose $s < s^*$. Then

$$\begin{aligned}
\mathbb{E}\left[(S[t]-s^*)J[t]|S[t]=s\right] &= (s-s^*)\mathbb{E}\left[J[t]|S[t]=s\right] \\
&\overset{(a)}{\leq} (s-s^*)g(s-s^*) \\
&\overset{(b)}{=} -|(s-s^*)g(s-s^*)|
\end{aligned}$$

  (a) uses $(s-s^*) < 0$ and (47); (b) uses $(s-s^*) < 0$ and (44).

It follows that for all $s \in [s_{min}, s_{max}]$ we have

$$\mathbb{E}\left[(S[t]-s^*)J[t]|S[t]=s\right] \leq -|(s-s^*)g(s-s^*)|$$

Thus

$$\mathbb{E}\left[(S[t]-s^*)J[t]|S[t]\right] \leq -\left|(S[t]-s^*)g(S[t]-s^*)\right|$$

Taking expectations of both sides of the above inequality and using the law of iterated expectations yields

$$\mathbb{E}\left[(S[t]-s^*)^2\right] \leq -\mathbb{E}\left[|(S[t]-s^*)g(S[t]-s^*)|\right]$$

Substituting the above inequality into (51) gives

$$\frac{1}{2}\mathbb{E}\left[(S[t+1]-s^*)^2\right] - \frac{1}{2}\mathbb{E}\left[(S[t]-s^*)^2\right] \leq \frac{\epsilon^2 c}{2} - \epsilon\mathbb{E}\left[|(S[t]-s^*)g(S[t]-s^*)|\right]$$

Summing over all $t \in \{0, 1, \ldots, T-1\}$ and using telescoping sums on the left-hand-side gives:[14]

$$\frac{1}{2}\mathbb{E}\left[(S[t]-s^*)^2\right] - \frac{1}{2}(S[0]-s^*)^2 \leq \frac{T\epsilon^2 c}{2} - \epsilon\sum_{t=0}^{T-1}\mathbb{E}\left[|(S[t]-s^*)g(S[t]-s^*)|\right]$$

Rearranging terms proves (50). □

Note that if the assumptions of the above theorem hold for the piecewise linear bounding function $g(x)$ given in (48) then

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[|(S[t]-s^*)g(S[t]-s^*)|\right] \leq \beta\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[(S[t]-s^*)^2\right]$$

[14]The term *telescoping sum* is used for an add-subtract scenario of the type $\sum_{t=0}^{T-1}(h[t+1]-h(t)) = (\cancel{h(1)}-h(0))+(h(2)-\cancel{h(1)})+...+(\cancel{h(T-1)}-\cancel{h(T-2)})+(h(T)-\cancel{h(T-1)}) = h(T)-h(0)$.

where $\beta = \min\{\beta_1, \beta_2\}$. The expression $\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[(S[t]-s^*)^2\right]$ can be called the *mean square error* between the estimate $S[t]$ and the true value s^* as the algorithm runs over T slots. In this special case, the above theorem says that we can choose ϵ arbitrarily small to ensure the mean square error is at most $O(\epsilon)$, provided that we run the algorithm for a number of slots proportional to $1/\epsilon^2$. Note that a *variable stepsize* $\epsilon[t]$ can often be used to yield a faster convergence time, but such methods are not adaptive to cases when s^* can change over time.

F. Simulation

This subsection provides simulation results for the multi-access system using the estimation procedure (38)-(39) with the parameters $a[t], b[t], c[t]$ given in Section VII-C. We use $m = 200$ and $S[0] = 2$. To understand how the algorithm adapts to changes in the number of users, we partition the time horizon of T slots into three phases. The number of users is constant over each phase but changes abruptly in between phases. The algorithm does not know the phase structure or the number of users in each phase, it simply adapts to the observed idle/success/collision information. We assume:

- The number of users in the first phase of the simulation is $n = 67$.
- The number of users in the second phase of the simulation is $n = 42$.
- The number of users in the third phase of the simulation is $n = 91$.

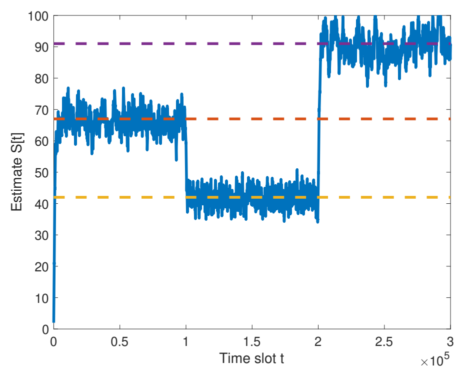

Fig. 11. Case 1: Estimation $S[t]$ versus time over the long timeline (3×10^5 slots) and with $\epsilon = 0.05$.

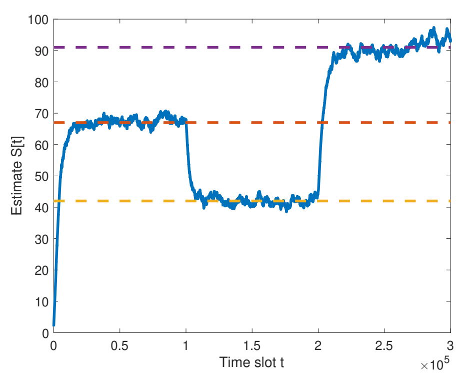

Fig. 12. Case 1: Estimation $S[t]$ versus time over the long timeline (3×10^5 slots) and with $\epsilon = 0.01$.

- Case 1 ($T = 3 \times 10^5$ slots): Fig. 11 and Fig. 12 show the results of $S[t]$ versus time for the case $\epsilon = 0.05$ and $\epsilon = 0.01$, respectively. The three dashed horizontal lines show the $n = 67$, $n = 42$, and $n = 91$ values which are optimal for the first, second, and third phases, respectively. As shown in the figures, the case $\epsilon = 0.05$ adapts more quickly to changes but does not settle into the optimal values as precisely as the case $\epsilon = 0.01$. The resulting average throughputs (averaged over the entire timeline) are:
  - $\epsilon = 0.05$: Throughput $\overline{X} = 0.3675$ (compare to $1/e = 0.3679$).
  - $\epsilon = 0.01$: Throughput $\overline{X} = 0.3668$ (compare to $1/e = 0.3679$).

  While these throughputs are very similar, the $\epsilon = 0.05$ simulation surprisingly has slightly higher total throughput. This is because, in the first phase, the $\epsilon = 0.01$ simulation takes a long time to initially converge from $S[0] = 2$ to the value $n = 67$ (shown as the red dashed horizontal line in Fig. 12), even though it eventually converges much more closely and accurately than the $\epsilon = 0.05$ simulation.

- Case 2 ($T = 3 \times 10^4$ slots): Here the simulation time is 10 times shorter in comparison to Case 1. We use the same phase structure but with 10 time smaller phases, so the algorithm has less time to adapt. Fig. 13 and Fig. 14 show results of $S[t]$ versus time for the case $\epsilon = 0.05$ and $\epsilon = 0.01$. It is evident from the figures that the large stepsize $\epsilon = 0.05$ is able to adapt reasonably well within these smaller timescales, while the $\epsilon = 0.01$ stepsize does not allow the algorithm to converge to the correct number of users before this number changes. The resulting average throughputs (averaged over the entire timeline) are significantly less than those of Case 1 because there is less time to adapt:
  - $\epsilon = 0.05$: Throughput $\overline{X} = 0.3626$ (compare to $1/e = 0.3679$).
  - $\epsilon = 0.01$: Throughput $\overline{X} = 0.3279$ (compare to $1/e = 0.3679$).

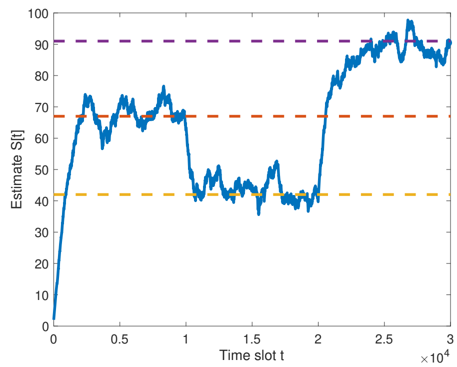

Fig. 13. Case 2: Estimation $S[t]$ versus time over the short timeline (3×10^4 slots) and with $\epsilon = 0.05$.

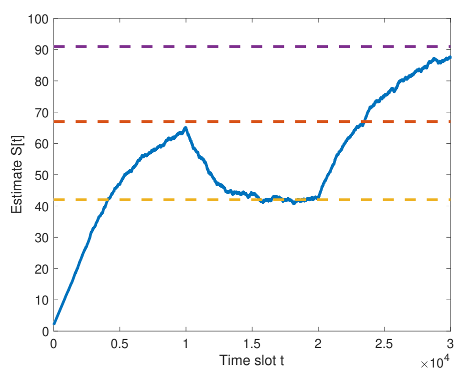

Fig. 14. Case 2: Estimation $S[t]$ versus time over the short timeline (3×10^4 slots) and with $\epsilon = 0.01$.

VIII. RANDOM ACCESS WITH RANDOMLY ARRIVING USERS (EACH WITH ONE PACKET)

Our random access analysis used a basic model with a fixed number of users n, all with an infinite number of packets to send. This section treats the opposite scenario where there are an infinite number of randomly arriving users, but all users have one and only one packet to send. They stay in the system only as long as it takes to successfully send the packet. Remarkably, our understanding of the fixed-user scenario lends significant insight into this new scenario. Further, the throughput values attained (for large n) under slotted Aloha, slotted CSMA, and slotted CSMA/CD are exactly the same as before.

The basic model is this:

- Fixed time slots $t \in \{0, 1, 2, \ldots\}$.
- Users arrive according to a (continuous time) Poisson arrival process of rate λ, where $0 < \lambda \leq 1$. The *memoryless property* of the Poisson arrival process means that the number of users A_d who arrive over an interval of time of size d is independent of system behavior before the start of the interval. Further, A_d has the $Poisson(\lambda d)$ probability mass function:
$$P[A_d = k] = \frac{(\lambda d)^k}{k!} e^{-\lambda d} \quad \forall k \in \{0, 1, 2, ...\} \tag{52}$$
  It is assumed that $\lambda \leq 1$ because the case $\lambda > 1$ cannot be stably supported since the maximum number of successes is one packet per slot.
- All users who arrive in the middle of a slot or mini-slot wait until the next slot boundary to start their activity.

A. Drift for discrete time Markov chains

The following drift result for Markov chains is useful. Suppose $\{Q(t)\}_{t=0}^{\infty}$ is time-homogenous discrete time Markov chain (DTMC) with positive integer state space $S \in \{0, 1, 2, \dots\}$ and transition probabilities:

$$P_{i,j} = P[Q(t+1) = j | Q(t) = i] \quad \forall i, j \in S$$

Define the *drift in state* n by

$$D_n = \mathbb{E}\left[Q(t+1) - Q(t) | Q(t) = n\right] \quad \forall n \in \{0, 1, 2, \dots\}$$

Lemma 2: (DTMC drift) Suppose the DTMC $\{Q(t)\}_{t=0}^{\infty}$ is irreducible, meaning that it is a path of nonzero probability between every two states $i, j \in \{0, 1, 2, \dots\}$. As a (non-crucial) technicality, also assume the DTMC is aperiodic.[15]

a) (Negative drift condition) Suppose there is an $\epsilon > 0$ and a threshold state $n^* \in \{0, 1, 2, \dots\}$ such that the following *negative drift condition* holds:

$$D_n \leq -\epsilon \quad \forall n \geq n^*$$

Further suppose $D_n < \infty$ for all $n \in \{0, 1, 2, \dots\}$. Then the DTMC is *stable*, meaning it has a stationary distribution.

b) (Positive drift condition) Suppose there is a threshold state $n^* \in \{0, 1, 2, \dots\}$ such that the following *positive drift condition* holds:

$$D_n > 0 \quad \forall n \geq n^*$$

Further suppose there is an integer $k > 0$ such that

$$Q(t+1) - Q(t) \geq -k \quad \forall t \in \{0, 1, 2, \dots\}$$

so that the DTMC cannot decrease by more than k from one slot to the next. Then the DTMC is *unstable*, meaning it does not have a stationary distribution. In particular

$$\lim_{t \to \infty} P[Q(t) \geq q] = 1 \quad \forall q \in \{0, 1, 2, \dots\}$$

Proof: See Chapter 3 Appendix 3A.5 in [2]. □

B. Slotted Aloha for Poisson arrivals

The slotted Aloha protocol for Poisson arrivals is this:

- New users wait until the next slot and then transmit with probability 1:
  - If the new user is successful (because it was the only transmission) then it leaves the system.
  - If the new user is not successful (because at least one other user transmitted) then it joins the group of *queued users*.
- Queued users transmit independently each slot with probability p_n, where index n is equal to the current number of queued users. If a queued user is successful (because it was the only transmission) then it leaves the system.

Let $Q(t)$ be the number of queued users in the system at the start of slot t. It should be emphasized that the *queue* is only conceptual and consists of those distributed users in their various locations who have arrived to the system but have not yet successfully transmitted their packet. Also, it is difficult for these users to know the value of $Q(t)$. Therefore, the algorithm is easier to implement if $p_n = p$ for all n, where p is some fixed value that does not depend on n. Unfortunately, we show that using any fixed value p creates an *unstable system*. The system can be stabilized using p_n proportional to $1/n$. In practice, stability is typically maintained when the unknown $Q(t)$ is approximated using some reasonable estimation technique.

Using $d = 1$ in (52), the number of new arrivals $A(t)$ in slot t has PMF

$$P[A(t) = k] = \frac{\lambda^k}{k!} e^{-\lambda} \quad \forall k \in \{0, 1, 2, \dots\}$$

and so

$$\mathbb{E}\left[A(t)\right] = \sum_{k=0}^{\infty} k P[A(t) = k] = \lambda$$

First assume $p_n = p$ for all $n \in \{0, 1, 2, \dots\}$, where p is a fixed transmission probability that satisfies $0 < p < 1$. The process $\{Q(t)\}_{t=0}^{\infty}$ can be seen to be a *discrete time Markov chain (DTMC)* with state space $S = \{0, 1, 2, \dots\}$ and transition probabilities P_{ij} for $i, j \in S$ computed by considering the probability of having k new arrivals in a slot. Specifically

$$P_{0,0} = P[A(t) \in \{0, 1\} | Q(t) = 0] = e^{-\lambda} + \lambda e^{-\lambda}$$

[15] For a general DTMC, a stationary probability distribution $(\pi_i)_{i \in S}$ is a PMF that satisfies $\pi_j = \sum_{i \in S} \pi_i P_{ij}$ for all $j \in S$. A stationary distribution can exist even when the DTMC is periodic. For any irreducible DTMC with a finite or countably infinite state space, if a stationary distribution exists then it must be unique, and the time average fraction of time being in each state i converges to π_i (with probability 1) regardless of the intitial state $Q(0)$. If, in addition, the DTMC is aperiodic then it also holds that $\lim_{t \to \infty} P[Q(t) = i | Q(0) = q_0] = \pi_i$ for all $q_i \in S$.

Since it is impossible for the number of queued users to transition from 0 to 1 we have $P_{0,1} = 0$. For $j \geq 2$ we have

$$P_{0,j} = P[A(t) = j|Q(t) = 0] = \frac{\lambda^j}{j!}e^{-\lambda} \quad \forall j \in \{2, 3, 4, \ldots\}$$

Since it is impossible for the number of queued users to decrease by more than 1, we have

$$P_{i,j} = 0 \quad \text{whenever } j < i - 1$$

For $i \in \{1, 2, 3, \ldots\}$ we have

$$\begin{aligned} P_{i,i-1} &= P[A(t) = 0, \text{Exactly one queued user transmits}|Q(t) = i] \\ &= P[A(t) = 0|Q(t) = i]P[\text{Exactly one queued user transmits}|Q(t) = i] \\ &= e^{-\lambda} ip(1-p)^{i-1} \end{aligned}$$

The remaining P_{ij} values can be similarly obtained. However, it is more useful to consider the *drift*.

Define $S(t) \in \{0, 1\}$ as the number of successful users in slot t (the successful user can be either a new user or a queued user). Then

$$Q(t+1) - Q(t) = A(t) - S(t)$$

and the drift is

$$\begin{aligned} D_n &= \mathbb{E}\left[Q(t+1) - Q(t)|Q(t) = n\right] \\ &= \mathbb{E}\left[A(t)|Q(t) = n\right] - \mathbb{E}\left[S(t)|Q(t) = n\right] \\ &= \lambda - P_{succ}(n) \end{aligned}$$

where $P_{succ}(n) = P[S(t) = 1|Q(t) = n]$ is the success probability given $Q(t) = n$ and

$$P_{succ}(0) = P[A(t) = 1|Q(t) = 0] = \lambda e^{-\lambda}$$

and for $n \geq 1$:

$$\begin{aligned} P_{succ}(n) &= P[A(t) = 0, \text{one queued user transmits}|Q(t) = n] + P[A(t) = 1, \text{no queued user transmts}|Q(t) = n] \\ &= e^{-\lambda}np(1-p)^{n-1} + \lambda e^{-\lambda}(1-p)^n \end{aligned}$$

It follows that

$$\boxed{D_n = \lambda - e^{-\lambda}np(1-p)^{n-1} - \lambda e^{-\lambda}(1-p)^n \quad \forall n \geq 1}$$

and so, regardless of the value $p \in (0, 1)$ we have

$$\lim_{n\to\infty} D_n = \lambda$$

Since $\lambda > 0$, it follows that there is a threshold value n^* such that

$$D_n \geq \lambda/2 \quad \forall n \geq n^*$$

The DTMC drift lemma (Lemma 2) implies the DTMC is unstable! The system is also clearly unstable when $p = 0$ or $p = 1$. Therefore *Slotted Aloha with randomly arriving users is unstable for any fixed transmission probability* $p \in [0, 1]$.

This instability can be fixed by using a *dynamic transmission probability* p_n for $n \in \{1, 2, 3, \ldots\}$, so the drift becomes (compare to the previous boxed equality):

$$D_n = \lambda - e^{-\lambda}np_n(1-p_n)^{n-1} - \lambda e^{-\lambda}(1-p_n)^n \quad \forall n \geq 1$$

that is

$$\boxed{D_n = \lambda - (\lambda + np_n)e^{-\lambda}(1-p_n)^n \quad \forall n \geq 1}$$

Let $p_n = \min[\theta/n, 1]$ for $n \in \{1, 2, 3, \ldots\}$, where $\theta > 0$ is a value to be optimized. Then for $n \geq \theta$ we have $p_n = \theta/n$ and

$$D_n = \lambda - (\lambda + \theta)e^{-\lambda}(1 - \theta/n)^n \quad \forall n \geq \theta$$

and so

$$\lim_{n\to\infty} D_n = \lambda - (\lambda + \theta)e^{-(\lambda+\theta)}$$

The limit is strictly negative when

$$\lambda < (\lambda + \theta)e^{-(\lambda+\theta)}$$

The value of θ can be chosen to maximize the right-hand-side:

$$(\lambda+\theta)e^{-(\lambda+\theta)} \leq \max_{g>0}\{ge^{-g}\} = 1/e$$

where the maximum is attained when $g^* = 1$. We can therefore choose θ so that $\lambda + \theta = g^* = 1$. This is always possible because $\lambda \leq 1$. That is, we choose

$$\theta^* = 1 - \lambda$$

and so

$$p_n^* = \frac{1-\lambda}{n} \quad \forall n \in \{1, 2, 3, \ldots\}$$

which means $\lambda + np_n = 1$. In particular, substituting p_n^* into the above boxed equality yields

$$D_n = \lambda - e^{-\lambda}\left(1 - \frac{1-\lambda}{n}\right)^n \quad \forall n \geq 1$$

and so

$$\lim_{n\to\infty} D_n = \lambda - e^{-1} = -(1/e - \lambda)$$

Hence, if $\lambda < 1/e$, there is a threshold n^* such that

$$D_n \leq -(1/e - \lambda)/2 \quad \forall n \geq n^*$$

and the DTMC drift lemma (Lemma 2) implies the DTMC is stable. Thus, this slotted Aloha protocol, with dynamic transmission probability $p_n^* = (1-\lambda)/n$, stably supports any arrival rate λ that satisfies $\lambda < 1/e$. *Therefore, the maximum throughput is $1/e$, which is the same value as in the case of n users with infinitely many packets.*

C. Slotted CSMA for Poisson arrivals

Fix β with $0 < \beta < 1$. The slotted CSMA protocol for Poisson arrivals is this:[16] Activity takes place over variable-sized frames, where each frame ends with one mini-slot worth of idle (an idle frame has size β and a success or collision frame has size $1 + \beta$):

- New users are officially labeled as queued users. They wait until they hear a new frame (initiated after one mini-slot of idle) and then transmit as all queued users do.
- Queued users transmit independently at the start of each new frame with probability p_n, where index n is equal to the number of queued users at the start of the frame. If a queued user is successful (because it was the only transmission) then it leaves the system.

Let $\{Q(k)\}_{k=1}^{\infty}$ be the DTMC, where $Q(k)$ is the number of queued users at the start of frame k. As before we have

$$Q(k+1) - Q(k) = A(k) - S(k)$$

where $S(k) \in \{0, 1\}$ is the number of successful users on frame k, and $A(k)$ is the number of new arrivals. The drift is then

$$D_n = \mathbb{E}\left[A(k) - S(k)|Q(k) = n\right] = \mathbb{E}\left[A(k)|A(k) = n\right] - P_{succ}(n) \quad \forall n \in \{0, 1, 2, \ldots\}$$

where $P_{succ}(n)$ is the success probability for a frame k that has $Q(k) = n$. For $n \geq 1$ we have

$$P_{succ}(n) = np_n(1-p_n)^{n-1} \quad \forall n \geq 1$$

Also it can be shown for $n \geq 1$:

$$\begin{aligned}\mathbb{E}\left[A(k)|Q(k) = n\right] &= \lambda\beta P[Idle] + \lambda(1+\beta)(1 - P[Idle]) \\ &= \lambda\beta(1-p_n)^n + \lambda(1+\beta)(1-(1-p_n)^n)\end{aligned}$$

As before, define $p_n = \min[\theta/n, 1]$ for $n \in \{1, 2, 3, \ldots\}$. Then $p_n = \theta/n$ for $n \geq \theta$ and

$$D_n = \lambda\beta(1-\theta/n)^{n-1} + \lambda(1+\beta)(1-(1-\theta/n)^n) - \theta(1-\theta/n)^{n-1} \quad \forall n \geq \theta$$

Thus

$$\begin{aligned}\lim_{n\to\infty} D_n &= \lambda\beta e^{-\theta} + \lambda(1+\beta)(1-e^{-\theta}) - \theta e^{-\theta} \\ &= \lambda(1+\beta) - (\lambda+\theta)e^{-\theta}\end{aligned}$$

[16] A variation on this protocol has new users who arrive during a mini-slot to transmit with probability 1 at the start of the next frame, while new users who arrive at other times are added to the queued users. There is no difference in the resulting throughput and the protocol treated here (which adds all new users directly to the queue) is simpler to analyze.

This is negative whenever

$$\lambda(1+\beta) < (\lambda+\theta)e^{-\theta}$$

That is, we want

$$\lambda < \frac{\theta}{(1+\beta)e^{\theta}-1}$$

where the right-hand-side is, remarkably, identical to (29). The DTMC drift lemma (Lemma 2) ensures stability whenever λ is less than the right-hand-side. To maximize the right-hand-side, we can choose θ the same way as given in the analysis for n users, each with infinitely many packets to send: An approximate optimum is $\theta^* \approx \sqrt{2\beta}$ and $\lambda^* \approx 1/(1+\sqrt{2\beta})$.

D. Slotted CSMA/CD for Poisson arrivals

Here we consider the same protocol as slotted CSMA, but cut collisions short to 2β. Using a similar analysis gives

$$D_n = \mathbb{E}\left[A(k)|Q(k)=n\right] - P_{succ}(n)$$

where

$$P_{succ}(n) = np_n(1-p_n)^{n-1}$$

and

$$\begin{aligned}\mathbb{E}\left[A(k)|Q(k)=n\right] &= \lambda\beta P[Idle] + \lambda(1+\beta)P[Succ] + \lambda(2\beta)P[Collision]\\ &= \lambda\beta(1-p_n)^n + \lambda(1+\beta)np_n(1-p_n)^{n-1} + 2\lambda\beta(1-(1-p_n)^n - np_n(1-p_n)^{n-1})\end{aligned}$$

Using $p_n = \min[\theta/n, 1]$ gives for $n \to \infty$

$$\lim_{n\to\infty} D_n = \lambda\beta e^{-\theta} + \lambda(1+\beta)\theta e^{-\theta} + 2\lambda\beta(1-e^{-\theta}-\theta e^{-\theta}) - \theta e^{-\theta}$$

This is negative when

$$\lambda < \frac{\theta e^{-\theta}}{2\beta + \theta e^{-\theta} - \beta e^{-\theta} - \beta\theta e^{-\theta}}$$

where the right-hand-side is identical to the result (30) for CSMA/CD in the context of n users with infinitely many packets to send. The optimal value $\theta^* = 0.7680$ given there can be used, yielding the same maximum throughput value $\lambda^* = 1/(1+3.31\beta)$. The protocol is stable when $\lambda < 1/(1+3.31\beta)$.

APPENDIX A – PROBABILITY DETAILS

Fix Ω as a nonempty set that shall be our sample space. If $A \subseteq \Omega$ we define the complement of A, written A^c by

$$A^c = \{\omega \in \Omega : \omega \notin A\}$$

Let ϕ denote the empty set, being the set with no elements. It holds that $\phi \subseteq \Omega$, $\phi^c = \Omega$, and $\Omega^c = \phi$.

A. The set $\mathcal{F}$ and the function P

Let $\mathcal{F}$ be a set of subsets of Ω. We often use the word *collection* when talking about a set of sets, so $\mathcal{F}$ can also be called a *collection* of subsets of Ω (the words "collection" and "set" have the same meaning here). We say that $\mathcal{F}$ is a *sigma algebra on* Ω if the following three properties hold:

- $\Omega \in \mathcal{F}$.
- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
- If $\{A_i\}_{i=1}^{\infty}$ satisfy $A_i \in \mathcal{F}$ for all $i \in \{1,2,3,\ldots\}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

These properties can be used to show that: $\phi \in \mathcal{F}$; the union of a finite or countably infinite number of sets in $\mathcal{F}$ is another set in $\mathcal{F}$; the intersection of a finite or countably infinite number of sets in $\mathcal{F}$ is another set in $\mathcal{F}$. The smallest sigma algebra on Ω is the 2-element set $\mathcal{F} = \{\phi, \Omega\}$. The largest sigma algebra on Ω is $\mathcal{F} = Pow(\Omega)$, where $Pow(\Omega)$ is the power set of Ω and is defined as the set of all subsets of Ω.

Given that $\mathcal{F}$ is a sigma algebra on Ω, we say that a function $P : \mathcal{F} \to \mathbb{R}$ is a *probability measure* if it satisfies the following three axioms of probability:

1) $P[A] \geq 0$ for all $A \in \mathcal{F}$.
2) $P[\Omega] = 1$.
3) If $\{A_i\}_{i=1}^{\infty}$ is a sequence of *disjoint* sets in $\mathcal{F}$, so that $A_i \cap A_j = \phi$ for all $i \neq j$, then

$$P[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$$

B. Consequences

A *probability space* is a triplet $(\Omega, \mathcal{F}, P)$ where Ω is a nonempty set called the *sample space*; $\mathcal{F}$ is a sigma algebra on Ω, and $P : \mathcal{F} \to \mathbb{R}$ is a probability measure. Sets in the collection $\mathcal{F}$ are called *events*. The axioms of probability can be used to prove that all events have probabilities that are numbers between 0 and 1:

$$0 \leq P[A] \leq 1 \quad \forall A \in \mathcal{F}$$

Hence, the function $P : \mathcal{F} \to \mathbb{R}$ can be viewed as a function of the form $P : \mathcal{F} \to [0, 1]$. It can also be proven that

$$P[A] + P[A^c] = 1 \quad \forall A \in \mathcal{F}$$

Finally, the probability axioms can be used to prove the following:

- Subset bound: If A and B are events that satisfy $A \subseteq B$, then $P[A] \leq P[B]$.
- Union bound: If $\{A_i\}_{i=1}^{\infty}$ are events then

$$P[\cup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} P[A_i]$$

C. A function $X : \Omega \to \mathbb{R}$ that is not a random variable

Fix a probability space $(\Omega, \mathcal{F}, P)$. A *random variable* is a function $X : \Omega \to \mathbb{R}$ that satisfies the following *measurability property*:

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R} \tag{53}$$

If $\mathcal{F} = Pow(\Omega)$ then all functions $X : \Omega \to \mathbb{R}$ are random variables because all subsets of Ω are in $\mathcal{F}$, so the measurability property (53) holds trivially. It is easy to construct an example function $X : \Omega \to \mathbb{R}$ that is *not* a random variable if we use a finite sample space Ω and a sigma algebra other than $Pow(\Omega)$. For example, consider

$$\begin{aligned} \Omega &= \{blue, red, green, black\} \\ \mathcal{F} &= \{\phi, \Omega, \{blue, red, green\}, \{black\}\} \end{aligned}$$

It can be shown that $\mathcal{F}$ is a sigma algebra on Ω, but $\mathcal{F}$ does not include all subsets of Ω. Define a function $X : \Omega \to \mathbb{R}$ by

$$\begin{aligned} X(blue) &= 0 \\ X(red) &= 0 \\ X(green) &= 1.3 \\ X(black) &= 7 \end{aligned}$$

Then

$$\{X \leq 1\} = \{blue, red\} \notin \mathcal{F}$$

and so the function $X : \Omega \to \mathbb{R}$ is not a random variable.

Of course, if we changed the sigma algebra to $\mathcal{F} = Pow(\Omega)$ then the same function $X : \Omega \to \mathbb{R}$ defined above would be a random variable.

D. Discrete sample spaces

If Ω is a finite or countably infinite set, we typically use $\mathcal{F} = Pow(\Omega)$, which means all subsets of Ω are events and all functions $X : \Omega \to \mathbb{R}$ are random variables. In particular, all 1-element subsets of Ω are events and hence $P[\{\omega\}]$ exists for all $\omega \in \Omega$. Then the probability measure $P : Pow(\Omega) \to [0, 1]$ must be

$$P[A] = \sum_{\omega \in A} P[\{\omega\}] \quad \forall A \subseteq \Omega \tag{54}$$

When Ω is an uncountably infinite set then we cannot define P according to (54) because summation is only defined when summing a finite or countably infinite number of values. Also, if Ω is an uncountably infinite set then we typically *do not* use $\mathcal{F} = Pow(\Omega)$ because it leads to surprising contradictions. For these uncountably infinite cases, we can typically use the best probability space in the world, defined in the next subsection.

E. The best probability space in the world

It can be shown that for any positive integer n, the multidimensional set $\mathbb{R}^n$ can be put into one-to-one correspondence with the (1-dimensional) unit interval $[0,1)$. Also, the set $\mathbb{R}^{\mathbb{N}}$ of all real-valued infinite sequences can be put into one-to-one correspondence with $[0,1)$. It turns out that most practical probability problems, including problems that have either countably or uncountably infinite sample spaces, can be reformulated to use the unit interval as the sample space, so that

$$\Omega = [0,1)$$

Specifically, consider the following probability space that we call the *best probability space in the world*:

$$(\Omega, \mathcal{F}, P) = ([0,1), \mathcal{B}([0,1)), \mu)$$

where:

- $\mathcal{B}([0,1))$ denotes the *Borel sigma algebra on* $[0,1)$, being the smallest sigma algebra on $[0,1)$ that includes the intervals $[0,x]$ for all $x \in [0,1)$.
- $\mu : \mathcal{B}([0,1)) \to [0,1]$ is the Borel measure on $[0,1)$ that maps every set in $\mathcal{B}([0,1))$ to its *length*.

In particular, it can be shown there is a unique measure $\mu : \mathcal{B}([0,1)) \to [0,1]$ such that

$$\mu([0,x]) = x \quad \forall x \in [0,1)$$

For this measure, it follows that for all a, b with $0 \leq a < b < 1$ we must have

$$\mu((a,b]) = b - a$$

and

$$\mu(\{x\}) = 0 \quad \forall x \in [0,1)$$

Therefore, for all a, b with $0 \leq a < b < 1$ we have

$$\begin{aligned} &\mu((a,b)) = \mu([a,b]) = \mu((a,b]) = \mu([a,b)) = b - a \\ &\mu((a,1)) = \mu([a,1)) = 1 - a \end{aligned}$$

It follows that if $A \subseteq [0,1)$ is a countable union of disjoint intervals then $\mu(A)$ is the sum of the interval lengths.

While $\mathcal{B}([0,1))$ does not contain *all* subsets of $[0,1)$, it has so many sets that, arguably, it contains *all subsets of* $[0,1)$ *of practical interest*.[17] For example, all finite or countably infinite subsets of $[0,1)$ are in $\mathcal{B}([0,1))$; all intervals of $[0,1)$ are in $\mathcal{B}([0,1))$; all open or closed subsets of $[0,1)$ are in $\mathcal{B}([0,1))$; all sets that can be formed by a finite or countably infinite procedure of complements, unions, and intersections of sets in $\mathcal{B}([0,1))$ are again in $\mathcal{B}([0,1))$.

On this probability space $([0,1), \mathcal{B}([0,1)), \mu)$, define the identity random variable $U : \Omega \to \mathbb{R}$ by $U(\omega) = \omega$ for all $\omega \in [0,1)$. It holds that U is uniformly distributed over $[0,1)$. In particular, $([0,1), \mathcal{B}([0,1)), \mu)$ is the simplest way to define a space that is big enough to contain a single random variable $U \sim \text{Unif}[0,1)$.

Arguably, any probability space that contains a single random variable $U \sim \text{Unif}[0,1)$ can be used to model any practical probability problem within the world of science, technology, and engineering. Why? Given $U \sim \text{Unif}[0,1)$, we can write U in its base-2 binary expansion:[18]

$$U = \sum_{i=1}^{\infty} U_i 2^{-i}$$

where $(U_1, U_2, U_3, \ldots)$ does not have an infinite tail of 1s. It can be shown that $\{U_i\}_{i=1}^{\infty}$ are i.i.d. Bernoulli(1/2) random variables, so

$$P[U_i = 0] = P[U_i = 1] = 1/2 \quad \forall i \in \{1, 2, 3, \ldots\}$$

By taking measurable functions of U (or of its bits $\{U_i\}_{i=1}^{\infty}$) we can generate an infinite sequence of mutually independent random variables $\{X_i\}_{i=1}^{\infty}$ where each X_i has any desired CDF F_{X_i}. Also by taking measurable functions of U, we can generate a countably infinite collection of mutually independent Brownian motion processes! In particular, a single random variable $U \sim \text{Unif}[0,1)$ is rich enough to generate the randomness needed for any practical computer simulation. This argument implicitly assumes that a "practical computer simulation" is one that uses a finite or countably infinite sequence of calls to a random number generator.

The probability space $([0,1), \mathcal{B}([0,1)), \mu)$ has its shortcomings: It cannot support an *uncountably infinite* number of mutually independent Bernoulli(1/2) random variables. There are larger probability spaces that can. However, the probability space $([0,1), \mathcal{B}([0,1)), \mu)$ is good enough for most applications.

[17] It is possible to construct a subset of $[0,1)$ that is not in $\mathcal{B}([0,1))$ by a deep thought experiment that uses a set theory axiom called the *axiom of choice* to make an uncountably infinite number of decisions. Such sets exist but do not arise in practice.

[18] We use $[0,1)$ instead of $[0,1]$ so that we can cleanly write the base-2 expansion. Aside from this, the probability spaces $([0,1], \mathcal{B}([0,1]), \mu)$ and $([0,1), \mathcal{B}([0,1)), \mu)$ have similar properties and either can be used when needed.

Appendix B – Extensions of renewal theory

A. Delayed renewal theorem

Claim (Delayed renewal theorem) Let $\{T_i\}_{i=2}^{\infty}$ be i.i.d. positive random variables with finite mean $\mathbb{E}[T] > 0$. Let T_1 be a positive random variable with any distribution and having any dependence on $\{T_i\}_{i=2}^{\infty}$. Let $N(t)$ count the number of arrivals up to time t (so the first arrival is at time T_1, the second is at time $T_1 + T_2$, and so on). Then

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[T]} \quad \text{almost surely}$$

Proof: Fix $t \geq T_1 + T_2$ (so that $N(t) \geq 2$). The time t must lie between two arrival times and so (the reader is encouraged to draw the picture similar to Fig. 1):

$$\textstyle\sum_{i=1}^{N(t)} T_i \leq t < \sum_{i=1}^{N(t)+1} T_i$$

Since $N(t) \geq 2$, we can divide the above inequality by $N(t)$ to obtain

$$\tfrac{T_1}{N(t)} + \left(\tfrac{N(t)-1}{N(t)}\right)\left(\tfrac{1}{N(t)-1}\textstyle\sum_{i=2}^{N(t)} T_i\right) \leq \tfrac{t}{N(t)} < \tfrac{T_1}{N(t)} + \tfrac{1}{N(t)}\textstyle\sum_{i=2}^{N(t)+1} T_i \tag{55}$$

Since $N(t) \to \infty$ surely, we have

$$\lim_{t\to\infty} \frac{T_1}{N(t)} = 0 \quad \text{surely}$$
$$\lim_{t\to\infty} \frac{N(t)-1}{N(t)} = 1 \quad \text{surely}$$
$$\lim_{t\to\infty} \frac{1}{N(t)-1}\sum_{i=2}^{N(t)} T_i = \mathbb{E}[T] \quad \text{almost surely}$$
$$\lim_{t\to\infty} \frac{1}{N(t)}\sum_{i=2}^{N(t)+1} T_i = \mathbb{E}[T] \quad \text{almost surely}$$

where the last two equalities hold by the LLN. Taking $t \to \infty$ in (55) and substituting these equalities gives the following "sandwich" result

$$\mathbb{E}[T] \leq \lim_{t\to\infty} \frac{t}{N(t)} \leq \mathbb{E}[T] \quad \text{almost surely}$$

Since $\mathbb{E}[T] > 0$, it follows that $\lim_{t\to\infty} N(t)/t = 1/\mathbb{E}[T]$ almost surely. $\Box$

Claim (Delayed renewal-reward theorem) Let $\{T_i\}_{i=2}^{\infty}$ be i.i.d. positive random variables with finite mean $\mathbb{E}[T] > 0$, let $\{G_i\}_{i=2}^{\infty}$ be i.i.d. random variables with finite mean $\mathbb{E}[G]$, let T_1 and G_1 be any random variables (where T_1 is positive). Then

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[G]}{\mathbb{E}[T]} \quad \text{almost surely}$$

Proof: For any $t \geq T_1 + T_2$ (so $N(t) \geq 2$) we have

$$\begin{aligned}
\frac{R(t)}{t} &= \frac{1}{t}\sum_{i=1}^{N(t)} G_i \\
&= \left(\frac{N(t)}{t}\right)\left(\frac{1}{N(t)}\sum_{i=1}^{N(t)} G_i\right) \\
&= \left(\frac{N(t)}{t}\right)\left(\frac{G_1}{N(t)} + \left(\frac{N(t)-1}{N(t)}\right)\left(\frac{1}{N(t)-1}\sum_{i=2}^{N(t)} G_i\right)\right)
\end{aligned}$$

Taking $t \to \infty$ and using the previous result $N(t)/t \to 1/\mathbb{E}[T]$ almost surely proves the result. $\Box$

B. Limiting expectations and the elementary renewal theorem

Claim: (Elementary renewal-reward theorem) Let $\{T_i\}_{i=1}^{\infty}$ be i.i.d. positive random variables with finite expectation $\mathbb{E}[T]$. Let $\{G_i\}_{i=1}^{\infty}$ be i.i.d. random variables with finite expectation $\mathbb{E}[G]$. Then

a) $\lim_{t\to\infty} \frac{\mathbb{E}[N(t)]}{t} = \frac{1}{\mathbb{E}[T]}$.

b) $\lim_{t\to\infty} \frac{\mathbb{E}[R(t)]}{t} = \frac{\mathbb{E}[G]}{\mathbb{E}[T]}$.

The proof is nontrivial and is omitted for brevity (see, for example, [1]).

C. Reward sprinkling

Consider now a system with i.i.d. inter-renewal times $\{T_i\}_{i=1}^{\infty}$ that are positive random variables with finite mean $\mathbb{E}[T] > 0$, i.i.d. rewards $\{G_i\}_{i=1}^{\infty}$ that are i.i.d. with finite mean $\mathbb{E}[G]$. Rather than assuming that the reward of frame i is earned in bulk at the end of the frame, assume that the total reward G_i can be sprinkled out in any manner during the course of the frame. This includes having some parts of the reward earned in pieces at distinct times during the middle of the frame, and/or having the reward accumulated continuously as a fluid. The resulting $R(t)$ function now can change at any time (not just at renewal times).

We need to treat the changes in $R(t)$ carefully. Assume there is an i.i.d. random sequence $\{L_i\}_{i=1}^{\infty}$ of nonnegative random variables with finite mean $\mathbb{E}[L]$, and an i.i.d. random sequence $\{H_i\}_{i=1}^{\infty}$ of nonnegative random variables with a finite mean $\mathbb{E}[H]$. The value L_i is a bound on the amount of negative rewards that can be sprinkled during frame i. The value H_i is a bound on the amount of positive rewards that can be sprinkled over frame i. The total reward $R(t)$ accumulated at time $t \geq 0$ is equal to the sum reward accumulated over the $N(t)$ complete frames (which is $\sum_{i=1}^{N(t)} G_i$) plus a (possibly negative) partial reward associated with the current frame. Thus

$$-L_{N(t)+1} + \sum_{i=1}^{N(t)} G_i \leq R(t) \leq H_{N(t)+1} + \sum_{i=1}^{N(t)} G_i \quad \forall t \geq 0 \tag{56}$$

- Example scenario 1: Suppose there are no negative rewards accumulated over any frame. Then $G_i \geq 0$ surely for all i. We can use $L_i = 0$ and $H_i = G_i$ for $i \in \{1, 2, 3, \ldots\}$.
- Example scenario 2: Suppose the worst-case negative reward that can be sprinkled is some positive constant M. We can use $L_i = M$ and $H_i = G_i + M$ for all $i \in \{1, 2, 3. \ldots\}$.

Claim: Under the reward-sprinkling model (56) as described above, we have

$$\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[G]}{\mathbb{E}[T]} \quad \text{almost surely}$$

Proof: Assuming $t > 0$ and dividing (56) by t gives

$$\frac{-L_{N(t)+1}}{t} + \left(\frac{N(t)}{t}\right) \frac{1}{N(t)} \sum_{i=1}^{N(t)} G_i \leq \frac{R(t)}{t} \leq \frac{H_{N(t)+1}}{t} + \left(\frac{N(t)}{t}\right) \frac{1}{N(t)} \sum_{i=1}^{N(t)} G_i \tag{57}$$

We already know $N(t) \to \infty$ surely and

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[T]} \quad \text{almost surely}$$

$$\lim_{t\to\infty} \frac{1}{N(t)} \sum_{i=1}^{N(t)} G_i = \mathbb{E}[G] \quad \text{almost surely}$$

Thus, taking $t \to \infty$ in (57) would prove the desired result *if* we can ensure:

$$\lim_{t\to\infty} \frac{L_{N(t)+1}}{t} = 0 \quad \text{almost surely} \tag{58}$$

$$\lim_{t\to\infty} \frac{H_{N(t)+1}}{t} = 0 \quad \text{almost surely} \tag{59}$$

To prove (58) and (59), observe that for all $t > 0$:

$$L_{N(t)+1} = \sum_{i=1}^{N(t)+1} L_i - \sum_{i=1}^{N(t)} L_i$$

Assuming $t > 0$ and dividing both sides by t gives

$$\frac{L_{N(t)+1}}{t} = \left(\left(\frac{N(t)+1}{t}\right) \frac{1}{N(t)+1} \sum_{i=1}^{N(t)+1} L_i\right) - \left(\frac{N(t)}{t} \frac{1}{N(t)} \sum_{i=1}^{N(t)} L_i\right)$$

We already know $N(t) \to \infty$ surely, and $N(t)/t \to 1/\mathbb{E}[T]$ almost surely. Hence, almost surely, the right-hand-side of the above equality converges to $\frac{\mathbb{E}[L]}{\mathbb{E}[T]} - \frac{\mathbb{E}[L]}{\mathbb{E}[T]} = 0$. Thus, (58) holds. The same argument shows (59) holds. □

The above theorem shows that this different manner of accumulating rewards does not change the limiting average reward per unit time. However, a pathological counter-example can be constructed when $\{L_i\}_{i=1}^{\infty}$ and $\{H_i\}_{i=1}^{\infty}$ are not i.i.d. with a finite mean: Suppose $T_i = 1$ for all i and $G_i = 0$ for all i. However, at the halfway point of each frame i we get a large reward $H_i = i^2$, later at the $3/4$ point of each frame i we get a large negative reward $-L_i = -i^2$. This indeed yields $G_i = H_i - L_i = 0$

for all i, but taking a limit as $n \to \infty$ over integers $n \in \{0, 1, 2, ...\}$ gives (noting that at time $n + \frac{1}{2}$ we are in the middle of frame $(n+1)$):

$$\lim_{n\to\infty} \frac{R(n+\frac{1}{2})}{n+\frac{1}{2}} = \lim_{n\to\infty} \frac{(n+1)^2}{n+\frac{1}{2}} = \infty \quad \text{surely}$$
$$\lim_{n\to\infty} \frac{R(n)}{n} = \lim_{n\to\infty} 0 = 0 \quad \text{surely}$$

In this example, $\lim_{t\to\infty} R(t)/t$ does not exist because $R(t_n)/t_n$ can have different limiting values over different subsequences of times t_n that satisfy $t_n \to \infty$.

Appendix C – Details on almost sure convergence

Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of random variables. Fix $c \in \mathbb{R}$. We say that $Y_n \to c$ *surely* if:

$$\lim_{n\to\infty} Y_n(\omega) = c \quad \forall \omega \in \Omega$$

We say that $Y_n \to c$ *almost surely* if:

$$P\left[\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = c\right] = 1$$

Convergence *almost surely* is also called *convergence with probability 1*. It is clear that

$$\text{Convergence surely} \implies \text{Convergence almost surely}$$

Theorem 5: $Y_n \to c$ almost surely if and only if for all $\epsilon > 0$ we have

$$\lim_{n\to\infty} P[\cup_{i=n}^{\infty}\{|Y_i - c| \geq \epsilon\}] = 0 \tag{60}$$

Proof: See Appendix D. □

A. Comparison to convergence in probability

Theorem 5 helps to understand the relationship between *convergence almost surely* and *convergence in probability*. We say that $Y_n \to c$ *in probability* if for all $\epsilon > 0$ we have

$$\lim_{n\to\infty} P[|Y_n - c| \geq \epsilon] = 0$$

Comparing this condition to (60), it is apparent that (60) is more stringent because for all $n \in \{1, 2, 3, \ldots\}$ and all $\epsilon > 0$ we have

$$\{|Y_n - c| \geq \epsilon\} \subseteq \cup_{i=n}^{\infty}\{|Y_i - c| \geq \epsilon\}$$

and so

$$P[|Y_n - c| \geq \epsilon] \leq P[\cup_{i=n}^{\infty}\{|Y_i - c| \geq \epsilon\}]$$

Taking a limit of both sides as $n \to \infty$ proves that

$$\text{Convergence almost surely} \implies \text{Convergence in probability}$$

B. Sufficient condition for almost sure convergence

Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of random variables. Fix $c \in \mathbb{R}$.

Lemma 3: If there is an $\alpha > 0$ such that

$$\sum_{i=1}^{\infty} \mathbb{E}\left[|Y_i - c|^{\alpha}\right] < \infty$$

then $Y_n \to c$ almost surely.

Proof: Fix $\epsilon > 0$. By Theorem 5 it suffices to show

$$\lim_{n\to\infty} P[\cup_{i=n}^{\infty}\{|Y_i - c| \geq \epsilon\}] = 0$$

We have by the union bound

$$\begin{aligned} P[\cup_{i=n}^{\infty}\{|Y_i - c| \geq \epsilon\}] &\leq \sum_{i=n}^{\infty} P[|Y_i - c| \geq \epsilon] \\ &= \sum_{i=n}^{\infty} P[|Y_i - c|^{\alpha} \geq \epsilon^{\alpha}] \\ &\leq \sum_{i=n}^{\infty} \frac{\mathbb{E}\left[|Y_i - c|^{\alpha}\right]}{\epsilon^{\alpha}} \end{aligned}$$

where the final inequality holds by the Markov inequality. Taking a limit as $n \to \infty$ gives

$$\lim_{n\to\infty} P[\cup_{i=n}^{\infty}\{|Y_i - c| \geq \epsilon\}] \leq \frac{1}{\epsilon^{\alpha}}\left(\lim_{n\to\infty}\sum_{i=n}^{\infty}\mathbb{E}\left[|Y_i - c|^{\alpha}\right]\right) \tag{61}$$

Since $\sum_{i=1}^{\infty}\mathbb{E}\left[|Y_i - c|^{\alpha}\right] < \infty$, the limit of its tail sum is zero:

$$\lim_{n\to\infty}\sum_{i=n}^{\infty}\mathbb{E}\left[|Y_i - c|^{\alpha}\right] = 0$$

Substituting this limit into (61) gives the result. □

C. Proof of law of large numbers for finite variance case

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. random variables with finite mean μ and variance σ^2. For each $n \in \{1, 2, 3, \dots\}$ define

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

It follows by linearity of expectation and basic properties of the variance for a linear combination of independent random variables that

$$\begin{aligned} \mathbb{E}\left[M_n\right] &= \mu \quad \forall n \in \{1, 2, 3, \dots\} \\ Var(M_n) = \mathbb{E}\left[(M_n - \mu)^2\right] &= \frac{\sigma^2}{n} \quad \forall n \in \{1, 2, 3, \dots\} \end{aligned}$$

We first show almost sure convergence to μ when we sample only over indices that are perfect squares, that is, we consider $M_1, M_4, M_9, M_{16}, M_{25}$ and so on.

Lemma 4: $M_{n^2} \to \mu$ almost surely.

Proof: We use Lemma 3 with $\alpha = 2$:

$$\begin{aligned} \sum_{n=1}^{\infty}\mathbb{E}\left[(M_{n^2} - \mu)^2\right] &= \sum_{n=1}^{\infty} Var(M_{n^2}) \\ &= \sum_{n=1}^{\infty}\frac{\sigma^2}{n^2} \\ &< \infty \end{aligned}$$

□

We now use the fact that we have almost sure convergence over perfect squares to get almost sure convergence.

Lemma 5: $M_n \to \mu$ almost surely.

Proof: First suppose that, in addition to $\{X_i\}_{i=1}^{\infty}$ being i.i.d. with finite mean and variance, we have that all X_i are surely nonnegative (so $X_i(\omega) \geq 0$ for all $\omega \in \Omega$ and all $i \in \{1, 2, 3, \dots\}$). We prove the result in this special nonnegative case, then show that this special case can be used to prove the general case.

For each positive integer n, define k_n as the largest integer such that $k_n^2 \leq n$. Then

$$k_n^2 \leq n < (k_n + 1)^2 = k_n^2 + 2k_n + 1$$

It is clear that $\lim_{n\to\infty} k_n = \infty$. It is not difficult to show that

$$\lim_{n\to\infty}\frac{k_n^2}{n} = \lim_{n\to\infty}\frac{(k_n+1)^2}{n} = 1 \tag{62}$$

Since $X_i \geq 0$ surely for all i, we have for each positive integer n that

$$\sum_{i=1}^{k_n^2} X_i \leq \sum_{i=1}^{n} X_i \leq \sum_{i=1}^{(k_n+1)^2} X_i$$

Dividing both sides by n gives

$$\left(\frac{k_n^2}{n}\right)\left(\frac{1}{k_n^2}\sum_{i=1}^{k_n^2} X_i\right) \leq M_n \leq \left(\frac{(k_n+1)^2}{n}\right)\left(\frac{1}{(k_n+1)^2}\sum_{i=1}^{(k_n+1)^2} X_i\right)$$

That is

$$\left(\frac{k_n^2}{n}\right) M_{k_n^2} \leq M_n \leq \left(\frac{(k_n+1)^2}{n}\right) M_{(k_n+1)^2} \tag{63}$$

We already know from Lemma 4 that

$$\begin{aligned} M_{k_n^2} &\to \mu \quad \text{almost surely} \\ M_{(k_n+1)^2} &\to \mu \quad \text{almost surely} \end{aligned}$$

Taking a limit of (63) and using these results and (62) gives

$$\mu \leq \lim_{n\to\infty} M_n \leq \mu \quad \text{almost surely}$$

This proves the desired result for the special case when X_i are surely nonnegative.

For the general case when X_i can be negative, break X_i into its positive and negative parts:

$$X_i = X_i^+ - X_i^- \quad \forall i \in \{1, 2, 3, \ldots\}$$

where $X_i^+ = \max\{X_i, 0\}$ and $X_i^- = \max\{-X_i, 0\}$. Note that since $\{X_i\}_{i=1}^{\infty}$ are i.i.d., it holds that the nonnegative random variables $\{X_i^+\}_{i=1}^{\infty}$ are i.i.d., as are the nonnegative random variables $\{X_i^-\}_{i=1}^{\infty}$. It also holds that X_i^+ has a finite mean and variance, as does X_i^-. Define $X = X_1$ and note that

$$\mathbb{E}[X] = \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right]$$

Our proven result for nonnegative i.i.d. random variables with finite mean and variance implies:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} X_i^+ &\to \mathbb{E}\left[X^+\right] \quad \text{almost surely} \\ \frac{1}{n}\sum_{i=1}^{n} X_i^- &\to \mathbb{E}\left[X^-\right] \quad \text{almost surely} \end{aligned}$$

It follows that we almost surely have:

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i &= \left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^+\right) - \left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i^-\right) \\ &= \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right] \\ &= \mathbb{E}[X] \end{aligned}$$

□

Appendix D – Proof of Theorem 5

Without loss of generality we assume $c = 0$. Let $\mathbb{N} = \{1, 2, 3, ...\}$. Let $\{Y_n\}_{n=1}^{\infty}$ be random variables. Define

$$A = \left\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = 0\right\}$$

Observe that

$$\boxed{(\omega \in A) \iff \text{“For all } k \in \mathbb{N}\text{, there is an } n \in \mathbb{N} \text{ such that } |Y_i(\omega)| < 1/k \text{ for all } i \geq n.\text{”}}$$

Hence

$$\boxed{A = \cap_{k\in\mathbb{N}} \cup_{n\in\mathbb{N}} \cap_{i=n}^{\infty} \{|Y_i| < 1/k\}} \tag{64}$$

where

- $\cap_{k\in\mathbb{N}}$ means "For all $k \in \mathbb{N}$"
- $\cup_{n\in\mathbb{N}}$ means "there is an $n \in \mathbb{N}$"
- $\cap_{i=n}^{\infty}\{|Y_i| \leq 1/k\}$ means "$|Y_i| < 1/k$ for all $i \geq n$."

The representation (64) formally proves the set A is an event, that is, $A \in \mathcal{F}$. That is because $|Y_i|$ is a random variable (it is a continuous function of Y_i) and so $\{|Y_i| < 1/k\} \in \mathcal{F}$ for all $i, k \in \mathbb{N}$. Therefore $\cap_{i=n}^{\infty}\{|Y_i| < 1/k\}$ is a countable intersection of events and so $\cap_{i=n}^{\infty}\{|Y_i| < 1/k\} \in \mathcal{F}$ for all $n, k \in \mathbb{N}$. Since the countable union of events in $\mathcal{F}$ is also in $\mathcal{F}$, we obtain $\cup_{n\in\mathbb{N}} \cap_{i=n}^{\infty} \{|Y_i| < 1/k\} \in \mathcal{F}$ for all $k \in \mathbb{N}$, and the countable intersection of these events over all $k \in \mathbb{N}$ is also in $\mathcal{F}$.

Clearly $P[A] = 1$ if and only if $P[A^c] = 0$. By DeMorgan's law:

$$A^c = \cup_{k\in\mathbb{N}} \cap_{n\in\mathbb{N}} \cup_{i=n}^{\infty}\{|Y_i| \geq 1/k\}$$

For $k, n \in \mathbb{N}$ define

$$B_{n,k} = \cup_{i=n}^{\infty}\{|Y_i| \geq 1/k\}$$

It holds that for each fixed $k \in \mathbb{N}$ we have

$$B_{n,k} \searrow \cap_{n=1}^{\infty} B_{n,k}$$

and by continuity of probability

$$\lim_{n\to\infty} P[B_{n,k}] = P[\cap_{n=1}^{\infty} B_{n,k}] \quad \forall k \in \mathbb{N} \tag{65}$$

We want to show the following statement (which is equivalent to the statement of Theorem 5):

$$\left(\lim_{n\to\infty} P[B_{n,k}] = 0 \quad \forall k \in \mathbb{N}\right) \iff P[A^c] = 0 \tag{66}$$

To prove the forward direction of (66), suppose for all $k \in \mathbb{N}$ we have $\lim_{n\to\infty} P[B_{n,k}] = 0$. Fix $k \in \mathbb{N}$. By (65) we have

$$P[\cap_{n=1}^{\infty} B_{n,k}] = 0$$

This holds for all $k \in \mathbb{N}$ and so by Theorem 2

$$P[\cup_{k=1}^{\infty} \cap_{n=1}^{\infty} B_{n,k}] = 0$$

That is, $P[A^c] = 0$. This proves the forward direction of (66).

To prove the reverse direction, suppose $P[A^c] = 0$. Fix $k \in \mathbb{N}$. We have

$$A^c \supseteq \cap_{n\in\mathbb{N}} \cup_{i=n}^{\infty} \{|Y_i| \geq 1/k\}$$

Since $P[A^c] = 0$, it follows that

$$P[\cap_{n\in\mathbb{N}} \cup_{i=n}^{\infty} \{|Y_i| \geq 1/k\}] = 0$$

That is

$$P[\cap_{n\in\mathbb{N}} B_{n,k}] = 0$$

which by (65) gives

$$\lim_{n\to\infty} P[B_{n,k}] = 0$$

This holds for all $k \in \mathbb{N}$, which proves the reverse direction of (66).

Appendix E — Conditioning, expectation, and Bernoulli trials

Fix a probability space $(\Omega, \mathcal{F}, P)$.

A. *Conditioning*

If A and B are events with $P[B] > 0$ then the *conditional probability of A given B* is defined

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

We say that a sequence of events $\{B_i\}_{i=1}^{\infty}$ *partition the sample space* if they are disjoint and their union is Ω:

$$B_i \cap B_i = \phi \quad \forall i \neq j$$
$$\cup_{i=1}^{\infty} B_i = \Omega$$

The *law of total probability* states that if A is an event and $\{B_i\}_{i=1}^{\infty}$ are events that partition the sample space then

$$\boxed{P[A] = \sum_{i=1}^{\infty} P[A \cap B_i] = \sum_{i:P[B_i]>0} P[A|B_i]P[B_i]}$$

This formula is proven by observing

$$A = A \cap \Omega = A \cap (\cup_{i=1}^{\infty} B_i) = \cup_{i=1}^{\infty}(A \cap B_i)$$

and so

$$P[A] = P\left[\cup_{i=1}^{\infty}(A \cap B_i)\right]$$

Now apply the third axiom of probability to the sequence of disjoint events $\{A \cap B_i\}_{i=1}^{\infty}$.

B. Indicator random variable

Let A be an event. Define the binary-valued random variable 1_A by

$$1_A = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{else} \end{cases}$$

In particular, $1_A(\omega) = 1$ if and only if $\omega \in A$. The random variable 1_A is called an *indicator random variable* or *indicator function*.

C. Expectation

If X is a random variable, the *expectation of* X, written $\mathbb{E}[X]$, is a value that satisfies

$$\mathbb{E}[X] \in \mathbb{R} \cup \{\infty\} \cup \{-\infty\} \cup \{DNE\}$$

where DNE represents *does not exist*. We say the expectation is finite if $\mathbb{E}[X] \in \mathbb{R}$. The expectation $\mathbb{E}[X]$ depends only on the CDF of X, so all random variables with the same CDF have the same expectation. It can be shown that if X is a nonnegative random variable, meaning that $X(\omega) \geq 0$ for all $\omega \in \Omega$, then the expectation always exists and $\mathbb{E}[X] \in [0, \infty) \cup \{\infty\}$. It can further be shown that if X is *any* random variable (possibly one that can take negative values), then $\mathbb{E}[X] \in \mathbb{R}$ if and only if $\mathbb{E}[|X|] < \infty$.

Let a, b be real numbers, let A be an event, and let X, Y be random variables that have finite expectations. The expectation has the following important properties:

1) Expectation of indicator:
$$\mathbb{E}[1_A] = P[A]$$
2) Linearity of expectation:
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

In particular, linearity of expectation says that if X and Y have finite expectations then $aX + bY$ also has a finite expectation, and this expectation is given by the above formula. By induction, linearity of expectation can be used to prove that if n is a positive integer, $a_1, \ldots, a_n$ are real numbers, and $X_1, \ldots, X_n$ are random variables with finite expectations, then

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

This can be used, together with the indicator function property, to prove a useful formula for random variables that can take at most finitely many values. Fix n as a positive integer and let $\{x_1, \ldots, x_n\}$ be a set of n distinct real numbers. Let $X : \Omega \to \{x_1, ..., x_n\}$ be a random variable. Observe that for each $i \in \{1, ..., n\}$ we have $\{X = x_i\}$ is an event and

$$X = \sum_{i=1}^{n} x_i 1_{\{X = x_i\}}$$

Taking expectations of the above equality (and using linearity of expectation and the indicator function property) yields

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{E}\left[1_{\{X=x_i\}}\right] = \sum_{i=1}^{n} x_i P[X = x_i]$$

This formula can be generalized in certain special cases when $\mathbb{E}[|X|] < \infty$:

- If X is a *discrete random variable*, meaning $X : \Omega \to D$ where $D \subseteq \mathbb{R}$ is a nonempty set that is either finite or countably infinite, and if $\mathbb{E}[|X|] < \infty$, it can be shown that
$$\mathbb{E}[X] = \sum_{x \in D} x P[X = x]$$
- If X is a *continuous random variable* with a valid *probability density function* (PDF), and if $\mathbb{E}[|X|] < \infty$, then it can be shown
$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

It can be shown that if X is any nonnegative random variable then:

$$\mathbb{E}[X] = \int_0^{\infty} P[X > x] dx$$

with the understanding that $\mathbb{E}[X] = \infty$ if and only if the right-hand-side of the above equality is ∞. A formal derivation of this uses the Lebesgue integral and the *Fubini-Tonelli theorem* of measure theory:

$$\left(X = \int_0^{\infty} 1_{\{X > x\}} dx\right) \implies \left(\mathbb{E}[X] = \int_0^{\infty} \mathbb{E}\left[1_{\{X > x\}}\right] dx\right)$$

D. Variance

If X is a random variable with finite expectation $\mathbb{E}[X]$ we define the *variance of* X, written $Var(X)$, by

$$Var(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

It can be shown that

$$Var(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

The value $\mathbb{E}\left[X^2\right]$ is called the *second moment* of X. It can be shown that a random variable X has a finite second moment if and only if $\mathbb{E}[X]$ and $Var(X)$ are both finite. If X, Y are two random variables with finite second moments then

$$Var(X+Y) = Var(X) + Var(Y) - 2\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

We say that X and Y are *uncorrelated* if

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0$$

It holds that X and Y are uncorrelated if and only if

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$$

It can be shown that if X and Y are independent with finite expectations, then they must be uncorrelated.

If X is a random variable with finite variance, and if $a \in \mathbb{R}$, then

$$Var(aX) = a^2 Var(X)$$

If $\{X_1, ..., X_n\}$ are *pairwise uncorrelated* (meaning $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\,\mathbb{E}[X_j]$ for all $i \neq j$) then

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

E. Expectation inequalities

If X and Y are random variables with finite expectations then it can be shown

$$(X \leq Y) \implies \mathbb{E}[X] \leq \mathbb{E}[Y]$$

As a consequence, if X is a nonnegative random variable (so $X(\omega) \geq 0$ for all $\omega \in \Omega$) then for all $\epsilon > 0$ we have

$$X \geq \epsilon 1_{\{X \geq \epsilon\}}$$

and so

$$\mathbb{E}[X] \geq \epsilon \mathbb{E}\left[1_{\{X \geq \epsilon\}}\right]$$

which proves the *Markov inequality* for nonnegative random variables (for any $\epsilon > 0$):

$$P[X \geq \epsilon] \leq \frac{\mathbb{E}[X]}{\epsilon}$$

If X has mean μ and finite second moment then the Markov inequality implies the following *Chebyshev inequality*: For all $\epsilon > 0$ we have

$$P[|X - \mu| \geq \epsilon] = P[|X - \mu|^2 \geq \epsilon^2] \leq \frac{\mathbb{E}\left[|X - \mu|^2\right]}{\epsilon^2} = \frac{Var(X)}{\epsilon^2} \tag{67}$$

If X and Y are random variables with finite second moments $\mathbb{E}\left[X^2\right]$ and $\mathbb{E}\left[Y^2\right]$, then XY has finite expectation and

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[|XY|]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]$$

where the second inequality is a form of the *Cauchy-Schwarz inequality* for random variables X, Y with finite second moments:

$$\mathbb{E}[XY]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]$$

which can be proven by first assuming $0 < \mathbb{E}\left[Y^2\right] < \infty$, observing the function $f(a) = \mathbb{E}\left[(X - aY)^2\right]$ is always nonnegative, and then finding the minimizer $a^* \in \mathbb{R}$ (the case $\mathbb{E}\left[Y^2\right] = 0$ can be treated separately).

F. *Law of total expectation*

If X is a random variable with $\mathbb{E}[|X|] < \infty$ and if B is an event with $P[B] > 0$ then we define the *conditional expectation of X given B* by

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X1_B]}{P[B]}$$

It can be shown that this is the same as the expectation of a random variable Y that has distribution equal to the conditional distribution of X given B, that is

$$(P[Y \leq y] = P[X \leq y|B] \quad \forall y \in \mathbb{R}) \implies (\mathbb{E}[Y] = \mathbb{E}[X|B])$$

If X is a random variable with $\mathbb{E}[|X|] < \infty$ and if $\{B_i\}_{i=1}^{\infty}$ are events that partition the sample space then

$$X = \sum_{i=1}^{\infty} X1_{B_i}$$

and it can be shown that:[19]

$$\boxed{\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{E}[X1_{B_i}] = \sum_{i:P[B_i]>0} \mathbb{E}[X|B_i]\, P[B_i]}$$

This is called the *law of total expectation*.

Appendix F — Sequences of random variables

Let $\{X_i\}_{i=1}^{\infty}$ be an infinite sequence of random variables, each with finite mean and variance. It holds that

$$i.i.d. \implies \text{pairwise independent} \implies \text{pairwise uncorrelated}$$

Suppose $\{X_i\}_{i=1}^{\infty}$ are pairwise uncorrelated, possibly having different distributions, but all having the same mean and variance:

$$\mathbb{E}[X_i] = \mu \quad \forall i \in \{1, 2, 3, ...\}$$
$$Var(X_i) = \sigma^2 \quad \forall i \in \{1, 2, 3, ...\}$$

where $\mu \in \mathbb{R}$ and $0 < \sigma^2 < \infty$. For each positive integer n define

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$G_n = \frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^{n}(X_i - \mu)$$

Linearity of expectation and the variance properties for pairwise uncorrelated random variables can be used to show

$$\mathbb{E}[M_n] = \mu \quad \forall n \in \{1, 2, 3, ...\}$$
$$Var(M_n) = \frac{\sigma^2}{n} \quad \forall n \in \{1, 2, 3, ...\}$$

which motivates the *law of large numbers*, and

$$\mathbb{E}[G_n] = 0 \quad \forall n \in \{1, 2, 3, ...\}$$
$$Var(G_n) = 1 \quad \forall n \in \{1, 2, 3, ...\}$$

which motivates the *central limit theorem*. In particular, if we additionally assume $\{X_i\}_{i=1}^{\infty}$ are i.i.d. then the *central limit theorem* (CLT) implies that the distribution of G_n converges to the $N(0, 1)$ distribution, being the distribution of a Gaussian random variable with mean 0 and variance 1:

$$\lim_{n\to\infty} P[G_n \leq x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad \forall x \in \mathbb{R}$$

This type of convergence is called *convergence in distribution*. The *law of large numbers* (LLN) implies a stronger form of convergence: If $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with finite mean μ then $M_n \to \mu$ with probability 1.[20]

[19]The thing to be careful of here is whether or not we can push the expectation through the *infinite sum*. The Fubini-Tonelli theorem of measure theory ensures that $\mathbb{E}\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i=1}^{\infty} \mathbb{E}[X_i]$ whenever either X_i are nonnegative random variables for all i, or when $\sum_{i=1}^{\infty} \mathbb{E}[|X_i|] < \infty$. In the formula above we have $|X| = \sum_{i=1}^{\infty} |X1_{B_i}|$ and so $\mathbb{E}[|X|] < \infty$ implies $\sum_{i=1}^{\infty} \mathbb{E}\left[|X1_{B_i}|\right] < \infty$, which allows pushing the expectation through the infinite sum.

[20]While the standard statement of the LLN requires $\{X_i\}_{i=1}^{\infty}$ to be i.i.d. and to have finite mean μ (with possibly infinite variance), it can be shown that if $\{X_i\}_{i=1}^{\infty}$ are pairwise uncorrelated, have possibly different distributions with the same mean μ, and have bounded variances so that $Var(X_i) \leq \sigma^2_{max}$ for all $i \in \{1, 2, 3, \ldots\}$ for some $\sigma^2_{max} < \infty$, then we can again conclude $M_n \to \mu$ with probability 1.

Since $\mathbb{E}[M_n] = \mu$ and $Var(M_n)$ is known, we can fix $\epsilon > 0$ and use the *Chebyshev inequality* (67) to conclude

$$P[|M_n - \mu| \geq \epsilon] \leq \frac{Var(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \quad \forall n \in \{1, 2, 3, ...\} \tag{68}$$

A. Hoeffding inequality

A related *Hoeffding inequality* is useful when the random variables are bounded. Here we state the result without proof: Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of mutually independent random variables (they do not need to be identically distributed). Define $\mu_i = \mathbb{E}[X_i]$ for $i \in \{1, 2, 3, \ldots\}$. If there are real numbers a_i, b_i such that[21]

$$a_i \leq X_i(\omega) \leq b_i \quad \forall i \in \{1, 2, 3, \ldots\}, \quad \forall \omega \in \Omega$$

then

- Single-sided Hoeffding inequality: For all $\epsilon > 0$ we have

$$P\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \geq \epsilon\right] \leq \exp\left(-\frac{2\epsilon^2 n^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \quad \forall n \in \{1, 2, 3, \ldots\}$$

- Double-sided Hoeffding inequality: For all $\epsilon > 0$ we have

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i)\right| \geq \epsilon\right] \leq 2\exp\left(-\frac{2\epsilon^2 n^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \quad \forall n \in \{1, 2, 3, \ldots\} \tag{69}$$

In the special case when $\{X_i\}_{i=1}^{\infty}$ are i.i.d. with mean μ we have $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) = M_n - \mu$. If μ is unknown, but if there are known a, b values that satisfy $a \leq X_i \leq b$ surely for all i, the double-sided Hoeffding inequality can be used to find a 95% confidence interval $[M_n - \epsilon, M_n + \epsilon]$ based on n i.i.d. samples. That is, we can find a radius $\epsilon > 0$ such that

$$P[\mu \in [M_n - \epsilon, M_n + \epsilon]] \geq 0.95 \tag{70}$$

simply by finding an ϵ that makes the right-hand-side of (69) equal to 0.05. Also, the right-hand-side of (69) reduces to $2\exp(\frac{-2\epsilon^2 n}{(b-a)^2})$, which decreases exponentially fast with n (as compared to the $O(1/n)$ decrease of Chebyshev in (68)). Thus, the Hoeffding inequality is typically much tighter than the Chebyshev inequality. Of course, the assumption that all random variables X_i are in the bounded interval $[a, b]$ is more detailed than the assumption that $Var(X_i) < \infty$.

An important special case is when $\{X_i\}_{i=1}^{\infty}$ are i.i.d. and $Bern(p)$ with some unknown $p \in [0, 1]$ (see also Section III-F). In this case we can use $a = 0$, $b = 1$, and the right-hand-side of (69) becomes $2e^{-2\epsilon^2 n}$. In this special case we obtain the following radii ϵ for 95% confidence intervals $[M_n - \epsilon, M_n + \epsilon]$ for the unknown p:

$$\begin{aligned}
n = 100 &\implies \epsilon = 0.135810 \\
n = 200 &\implies \epsilon = 0.096032 \\
n = 300 &\implies \epsilon = 0.078410 \\
n = 10^3 &\implies \epsilon = 0.042947 \\
n = 10^4 &\implies \epsilon = 0.013581 \\
n = 10^5 &\implies \epsilon = 0.004295 \\
n = 10^6 &\implies \epsilon = 0.0013581
\end{aligned}$$

It should be emphasized that these radii ϵ are computed *before observing any of the random samples* $\{X_i\}$. The width of the confidence interval is 2ϵ, although the left-endpoint of the confidence interval is the random variable $M_n - \epsilon$.

The inequality (70) states that, even before observing the M_n random variable, we know the probability that the random interval $[M_n - \epsilon, M_n + \epsilon]$ includes the (unknown) value μ is greater than or equal to 0.95. In particular, inequality (70) holds *before looking at the value* M_n. One may wonder what the *conditional* probability is, given we observe M_n:

$$P[\mu \in [M_n - \epsilon, M_n + \epsilon]|M_n] = ??$$

The correct answer to this question is not particularly helpful: It depends on M_n, and it is always either 0 or 1 depending on the value M_n. That is because there is no randomness to the parameter μ: Either μ is in the given interval or it is not. For example, if we happen to observe $M_n = 0.87$ then:[22]

$$P[\mu \in [M_n - \epsilon, M_n + \epsilon]|M_n = 0.87] = P[\mu \in [0.87 - \epsilon, 0.87 + \epsilon]|M_n = 0.87] = \begin{cases} 1 & \text{if } \mu \in [0.87 - \epsilon, 0.87 + \epsilon] \\ 0 & \text{else} \end{cases}$$

[21] The assumption that X_i is surely in $[a, b]$ can be replaced by the assumption that X_i is almost surely in $[a, b]$, with no change in any probability result.

[22] A formal treatment of conditional probability given a random variable M_n recognizes that there can be many *versions* of $\mathbb{E}\left[1_{\{\mu \in [M_n - \epsilon, M_n + \epsilon]\}}|M_n\right]$, every version has the form $g(M_n)$ for some measurable function $g : \mathbb{R} \to \mathbb{R}$, and any two versions $g_1(M_n)$ and $g_2(M_n)$ satisfy $P[g_1(M_n) = g_2(M_n)] = 1$. Our consideration of the case $M_n = 0.87$ is consistent with the version given by $g(M_n)$ for $g(x) = 1_{\{\mu \in [x - \epsilon, x + \epsilon]\}}$ for all $x \in \mathbb{R}$.

and so if μ is unknown, the answer exists but cannot be evaluated. It is more useful to consider the unconditional probability in (70), which can be obtained from bounds such as the Chebyshev and Hoeffding inequalities.[23]

B. Relation to Gaussian distributions in statistics

In the case when $\{X_i\}_{i=1}^{\infty}$ are i.i.d. $N(\mu, \sigma^2)$ with unknown μ and σ^2 (with $\mu \in \mathbb{R}$ and $0 < \sigma^2 < \infty$) a remarkable result in the theory of statistics holds: A confidence interval I_n (with random endpoints and a random width that is roughly proportional to $1/\sqrt{n}$) can be constructed from n samples, using the sample mean and sample variance, such that $P[\mu \in I_n]$ can be *exactly* computed using a student's t-distribution [28]. Specifically, for any integer $n \geq 2$ and any real number $t > 0$ we can define the interval

$$I_n = \left[M_n - t\frac{\sqrt{V_n}}{\sqrt{n}}, M_n + t\frac{\sqrt{V_n}}{\sqrt{n}}\right] \tag{71}$$

where random variables M_n and V_n are the sample mean and sample variance

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$$

where dividing by $n-1$ in the sample variance, rather than n, ensures the result is unbiased. Defining the random variable T_n by[24]

$$T_n = \frac{\sqrt{n}}{\sqrt{V_n}}(M_n - \mu)$$

yields the remarkable result that T_n has a student's t-distribution with $n-1$ degrees of freedom, so $P[|T_n| > t]$ can be calculated for all $t > 0$. If we choose t such that $P[|T_n| > t] = 0.05$ then

$$\begin{aligned}
0.95 &= P[|T_n| \leq t] \\
&= P\left[\frac{\sqrt{n}}{\sqrt{V_n}}|M_n - \mu| \leq t\right] \\
&= P\left[|M_n - \mu| \leq t\frac{\sqrt{V_n}}{\sqrt{n}}\right] \\
&= P[\mu \in I_n]
\end{aligned}$$

The value $t > 0$ that yields $P[|T_n| > t] = 0.05$ is often called $t_{0.025,n-1}$ and is found from a lookup table to obtain the (one-sided) probability $P[T_n > t] = 0.025$. Symmetry of the PDF for a student's-t random variable with $n-1$ degrees of freedom implies $P[|T_n| > t] = 2P[T_n > t]$. The student's-$t$ CDF is continuous so $P[T_n = x] = 0$ for all $x \in \mathbb{R}$.

When the distribution of each X_i is not $N(\mu, \sigma^2)$, but when an upper bound on the variance is known, the Chebyshev inequality (68) can be used. The Chebyshev inequality is generally considered to be loose. For a tighter result when the random variables X_i are surely in some bounded interval $[a, b]$, the Hoeffding inequality (69) can be used. Observe that (68) and (69) are solid mathematical bounds that hold for any sample size $n \in \{1, 2, 3, \ldots\}$.

When the number of (non-Gaussian) samples n is large, an alternative statistical practice is to obtain *approximate bounds* and *approximate confidence intervals* via a conventional method of *batch means* that groups the i.i.d. samples into batches of sufficiently large size for the central limit theorem to suggest each batch mean is *approximately* Gaussian [28]. This can give *approximate* confidence intervals that are sometimes tighter than the (rigorous) confidence intervals obtained by Chebyshev or Hoeffding. The approximate confidence intervals are of the type (71) with the exception that the sample mean and sample variance are computed from the batch mean samples, not the raw samples. This method is particularly useful when the random variables X_i are not bounded, so that Hoeffding cannot be applied. In practice, the method of batch means often uses batches of size 15-20 samples each, although the size required for a good approximation to a Gaussian actually depends on the distribution of X_i.

[23]A review of the law of large numbers may be useful: Fix the (possibly unknown) distribution for X_1 with the (possibly unknown) mean μ. The inequality (70) implies that if we fix n and independently repeat the experiment to form an infinite i.i.d. sequence $\{M_{n,i}\}_{i=1}^{\infty}$, where each $M_{n,i}$ has the same distribution as our original M_n, then the law of large numbers guarantees that, as the number of experiments gets large, the fraction of experiments that result in $\mu \in [M_{n,i} - \epsilon, M_{n,i} + \epsilon]$ approaches a value that is at least 0.95. This can be directly tested by simulation when the distribution of X_1 (and the value μ) is known by the computer: Consider, say, 10^6 experiments. For each experiment $i \in \{1, \ldots, 10^6\}$ we independently generate n i.i.d. samples with the same distribution as X_1, call them $\{X_{n,i,j}\}_{j=1}^{n}$, and define $M_{n,i} = \frac{1}{n}\sum_{j=1}^{n} X_{n,i,j}$. The computer can then count how many experiments yield a randomly generated interval that contains μ. Of course, by independently repeating the experiment in this way, we view all random variables $X_{n,i,j} : \Omega \to \mathbb{R}$ as part of one larger experiment that takes place on a single probability space $(\Omega, \mathcal{F}, P)$, so a given outcome $\omega \in \Omega$ determines the values $X_{n,i,j}(\omega)$ for all $j \in \{1, \ldots, n\}, i \in \{1, 2, 3, \ldots\}$.

[24]Formally we define $T_n = 0$ if the (probability zero) event $V_n = 0$ holds (this can only hold if $X_i = X_1$ for all $i \in \{1, ..., n\}$).

REFERENCES

[1] S. Ross. *Introduction to Probability Models*. Academic Press, 8th edition, Dec. 2002.

[2] D. P. Bertsekas and R. Gallager. *Data Networks*. New Jersey: Prentice-Hall, Inc., 1992.

[3] J. Tsitsiklis. Analysis of a multiaccess control scheme. *IEEE Transactions on Automatic Control*, 32(11):1017–1020, 1987.

[4] S. Gollakota and D. Katabi. Zigzag decoding: combating hidden terminals in wireless networks. *SIGCOMM '08: Proceedings of the ACM SIGCOMM 2008 conference on Data communication*, 2008.

[5] A. S. Tehrani, A. G. Dimakis, and M. J. Neely. Sigsag: Iterative detection through soft message-passing. *IEEE Journal of Selected Topics in Signal Processing*, 5(8):1512 – 1523, Dec. 2011.

[6] Q. Zhao and L. Tong. A dynamic queue protocol for multiaccess wireless networks with multipacket reception. *IEEE Trans. on Wireless Comm.*, vol. 3, no. 6, pp. 2221-2231, Nov. 2004.

[7] J. Paek and M. J. Neely. Mathematical analysis of throughput bounds in random access with zigzag decoding. *Mobile Networks and Applications*, 16(2):255 – 266, 2011.

[8] E. Casini, R. De Gaudenzi, and O. Del Rio Herrero. Contention resolution diversity slotted aloha (crdsa): An enhanced random access schemefor satellite access packet networks. *IEEE Transactions on Wireless Communications*, 6(4):1408–1419, 2007.

[9] O. Herrero and Riccardo De Gaudenzi. A high-performance mac protocol for consumer broadband satellite systems. pages 512–512, 01 2009.

[10] Gianluigi Liva. Graph-based analysis and optimization of contention resolution diversity slotted aloha. *IEEE Transactions on Communications*, 59(2):477–487, 2010.

[11] Krishna R Narayanan and Henry D Pfister. Iterative collision resolution for slotted aloha: An optimal uncoordinated transmission policy. In *2012 7th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, pages 136–139. IEEE, 2012.

[12] Matteo Berioli, Giuseppe Cocco, Gianluigi Liva, and Andrea Munari. Modern random access protocols. *Foundations and Trends® in Networking*, 10(4):317–446, 2016.

[13] E. Paolini, G. Liva, and M. Chiani. Coded slotted aloha: A graph-based method for uncoordinated multiple access. *IEEE Transactions on Information Theory*, 61(12):6815–6832, 2015.

[14] A. Taghavi, A. Vem, J. Chamberland, and K. R. Narayanan. On the design of universal schemes for massive uncoordinated multiple access. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 345–349, 2016.

[15] A. Vem, K. R. Narayanan, J. Cheng, and J. Chamberland. A user-independent serial interference cancellation based coding scheme for the unsourced random access gaussian channel. In *2017 IEEE Information Theory Workshop (ITW)*, pages 121–125, 2017.

[16] K. K. G. Vishnu and K. R. Narayanan. Evaluation of interference-cancellation based mac protocol for vehicular communications. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pages 1–6, 2018.

[17] N. Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrscheinlichkeitstheorie verw. Geiete*, vol. 5:pp. 119–122, 1981.

[18] M. J. Neely. *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool, 2010.

[19] Bruce Hajek and Timothy Van Loon. Decentralized dynamic control of a multiaccess broadcast channel. *IEEE transactions on automatic control*, 27(3):559–569, 1982.

[20] Frank P Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(3):379–395, 1985.

[21] VA Mikhailov. Methods of random multiple access. *Candidate Eng. Thesis, Moscow Institute of Physics and Technology*, 1979.

[22] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[23] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[24] A. Nemirovski and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience Series in Discrete Mathematics, John Wiley, 1983.

[25] H. J. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.

[26] V. S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Springer, 2008.

[27] V. R. Joseph. Efficient Robbins-Monro procedure for binary data. *Biometrika*, 91(2):461–470, June 2004.

[28] A. Leon-Garcia. *Probability, Statistics, and Random Processes For Electrical Engineering, 3rd ed.* Pearson Education Inc., 2008.