Overview of Probability for Computational Biology and Bioinformatics

Lenwood S. Heath

Original References

- **Textbook:** Probability, Statistics, and Queueing Theory with Computer Science Applications, Arnold O. Allen

- **Additional reference:** Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Kishor S. Trivedi

Key Concepts in Probability

- **Sample, event, or probability space.** A set Ω.

- **Elementary event.** An element $\omega \in \Omega$.

- **An event.** A subset $A \subseteq \Omega$.

Axioms of a Probability Measure

A **probability measure** Pr is a real function on a family of events $\mathcal{F}$ that satisfies these axioms.

P1 If $A \in \mathcal{F}$, then $0 \leq \Pr[A]$.

P2 $\Pr[\Omega] = 1$.

P3 If $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$, then $\Pr[A \cup B] = \Pr[A] + \Pr[B]$.

P4 If $A_1, A_2, A_3, \ldots$ are such that $A_i \cap A_j = \emptyset$ when $i \neq j$, then

$$\Pr\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \Pr[A_n].$$

Interpreting the Axioms

P1 Probabilities are always nonnegative.

P2 The probability of the certain event is always 1.

P3 A probability measure is additive when applied to the union of two mutually exclusive events. We can use **P2** and **P3** to show that

$$\Pr[\bar{A}] \;=\; 1 - \Pr[A].$$

P4 A probability measure is additive when applied to the countable union of mutually exclusive events. No amount of induction will make **P3** imply **P4**. Axiom **P4** is vacuously true if $\mathcal{F}$ is finite.

Discrete Example: Coin Flips

- Sample space Ω consists of possible sequences of 3 coin flips. Elementary events:

$$\begin{aligned}
\omega_1 &= H, H, H \\
\omega_2 &= H, H, T \\
\omega_3 &= H, T, H \\
\omega_4 &= H, T, T \\
\omega_5 &= T, H, H \\
\omega_6 &= T, H, T \\
\omega_7 &= T, T, H \\
\omega_8 &= T, T, T
\end{aligned}$$

- $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$

Coin Flips Continued

- A probability measure Pr for a discrete sample space is determined by its values on the elementary events.

$$\begin{aligned}
\Pr[\omega_1] &= .1 \\
\Pr[\omega_2] &= .05 \\
\Pr[\omega_3] &= 0 \\
\Pr[\omega_4] &= .4 \\
\Pr[\omega_5] &= .15 \\
\Pr[\omega_6] &= .05 \\
\Pr[\omega_7] &= .10 \\
\Pr[\omega_8] &= .15
\end{aligned}$$

- It is easy to verify **P1–P4** for Pr. Hence Pr is indeed a probability measure.

Verifying the Axioms

For any $A \in \mathcal{F}$,

$$\Pr[A] = \sum_{\omega \in A} \Pr[\omega]. \tag{1}$$

Verify **P1**–**P4**:

P1 True since every $\Pr[\omega] \geq 0$.

P2 True since

$$\begin{aligned} \Pr[\Omega] &= \sum_{i=1}^{8} \Pr[\omega_i] \\ &= 1. \end{aligned}$$

P3 Follows from Equation 1.

P4 Vacuously true.

Notes on Coin Flips

- We have slightly simplified notation so that $\Pr[\omega_1]$ means $\Pr[\{\omega_1\}]$.

- Note that this Pr does not correspond to the probability measure that would arise from **independent** coin flips.

Continuous Example: Unit Interval

- Sample space Ω is the unit interval $[0, 1]$.

Unit Interval Continued

- To define a probability measure Pr in this case, it suffices to define it for the intervals:

$$\text{Pr}\,[a, b] \;=\; b - a.$$

- Verifying **P1–P4** requires an inductive proof.

- This is the continuous version of **uniform probability.**

- In the theory of integration, one writes

$$\text{Pr}\,[A] \;=\; \int_A 1.$$

  This integral is also known as the **Lebesgue measure** of the set A. It just formalizes our notion of the cumulative "length" of A. And it formalizes what it means that we can "integrate" over A.

More Notes on the Unit Interval

- We have again simplified notation so that $\Pr[a, b]$ means $\Pr[[a, b]]$.

- We do not actually do the verification of **P1**–**P4**. It would take some time to make a careful proof.

Random Sampling

- Can use uniform probability on a unit interval to define a **random real number** r **between 0 and 1.** The probability that r comes from any interval $[a, b]$ is proportional to the size of the interval:

$$\Pr[a \leq r \leq b] \;=\; b - a.$$

  That is, r is just as likely to be $\leq 1/2$ as it is $\geq 1/2$.

- Cannot use uniform probability to define a **random real number** r **between 0 and** ∞**.** Measure of the interval $[0, \infty)$ is infinite, not 1. Need to assign to an interval $[a, b]$ "near infinity" a probability "near 0."

Notes on Random Sampling

- Once again we have expanded our notation by using $\Pr[a \le r \le b]$ to mean $\Pr[a, b]$.

- Since

$$\begin{aligned} \Pr[r = a] &= \Pr[a \le r \le a] \\ &= 0, \end{aligned}$$

the probability that any particular point in the unit interval is sampled is zero.

Random Sample Exercise

- Sample space $\Omega = \{1, 2, \ldots, 1000\}$

- Uniform probability measure:

$$\Pr[A] \;=\; \frac{|A|}{1000}$$

  for all $A \subseteq \Omega$

- **EXERCISE.** Suppose we want to know the probability that a randomly sampled element of Ω is prime.

  – What is the event A that corresponds to such a sample?

  – What is $\Pr[A]$?

Conditional Probability

Normal setup: probability space Ω; events are subsets $A, B, C, \ldots \subseteq \Omega$; probability measure Pr. **Useful new concepts:**

- **Conditional probability:**

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

  Always assume $\Pr[B] > 0$.

- **Independent events:**

$$\Pr[A \cap B] = \Pr[A] \Pr[B]$$

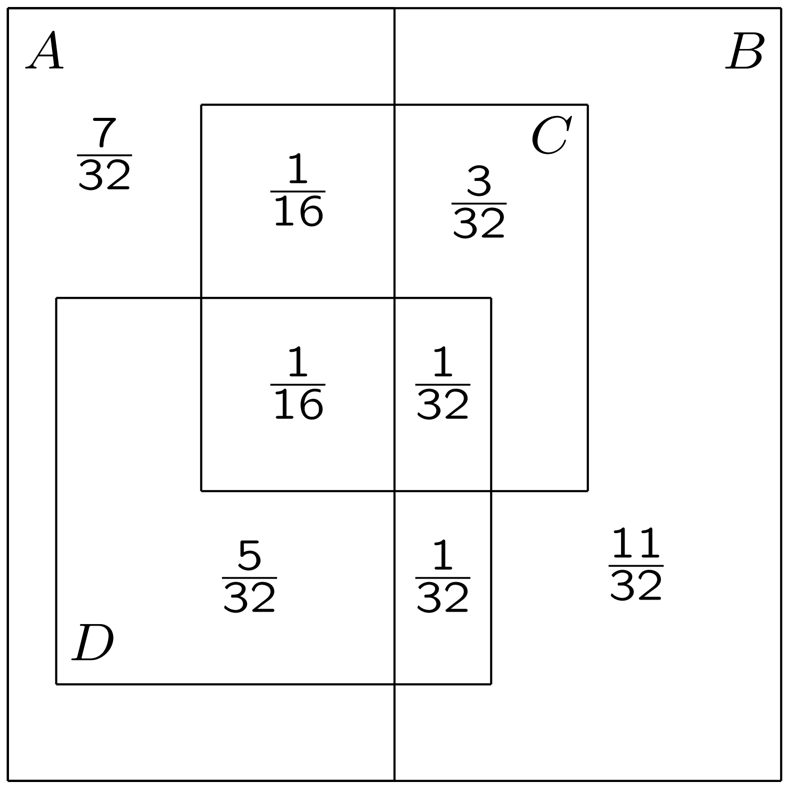
- **Disjoint events:**

$$A \cap B = \emptyset,$$

  so

$$\Pr[A \cap B] = 0.$$

Example



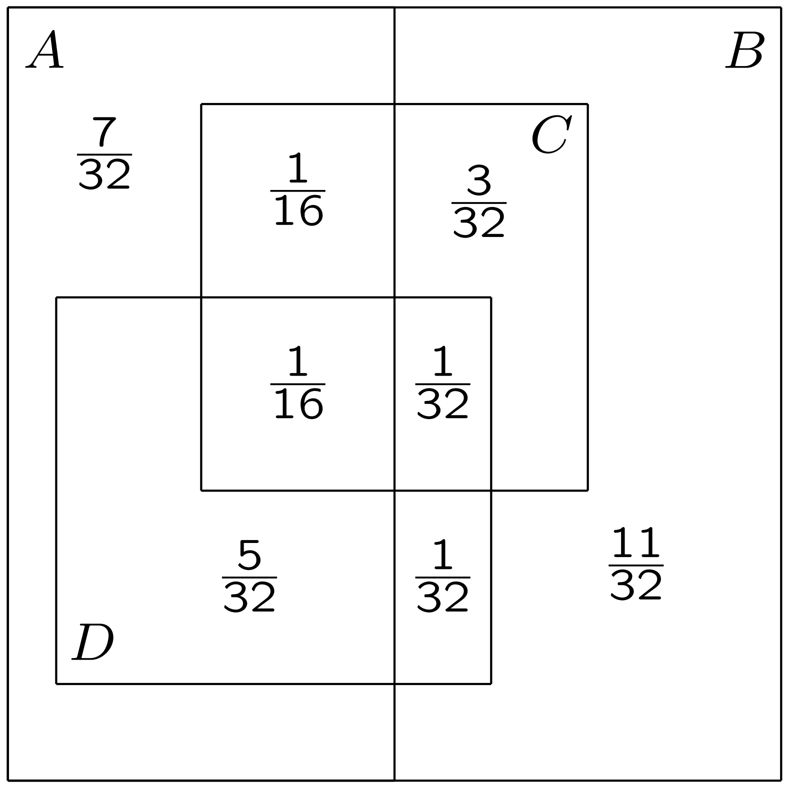
$$\Pr[A] = \frac{1}{2}$$

$$\Pr[B] = \frac{1}{2}$$

$$\Pr[C] = \frac{1}{4}$$

$$\Pr[D] = \frac{9}{32}$$

Conditional Probabilities



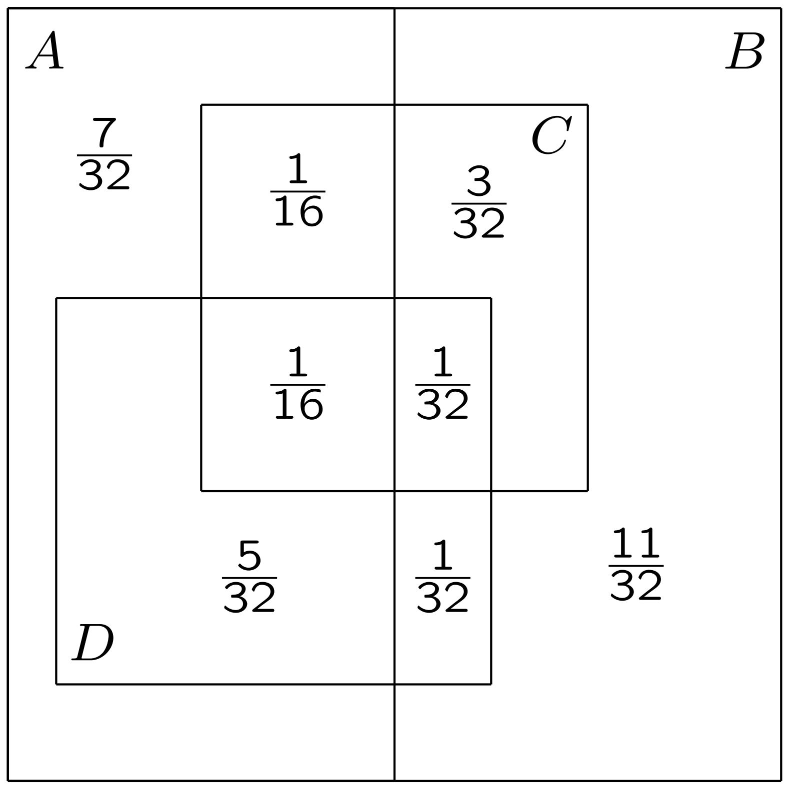
$$\Pr[C \mid A] = \frac{1}{4}$$

$$\Pr[B \mid C] = \boxed{\quad ?}$$

$$\Pr[D \mid C] = \boxed{\quad ?}$$

$$\Pr[D \mid \bar{C}] = \boxed{\quad ?}$$

Independent Events

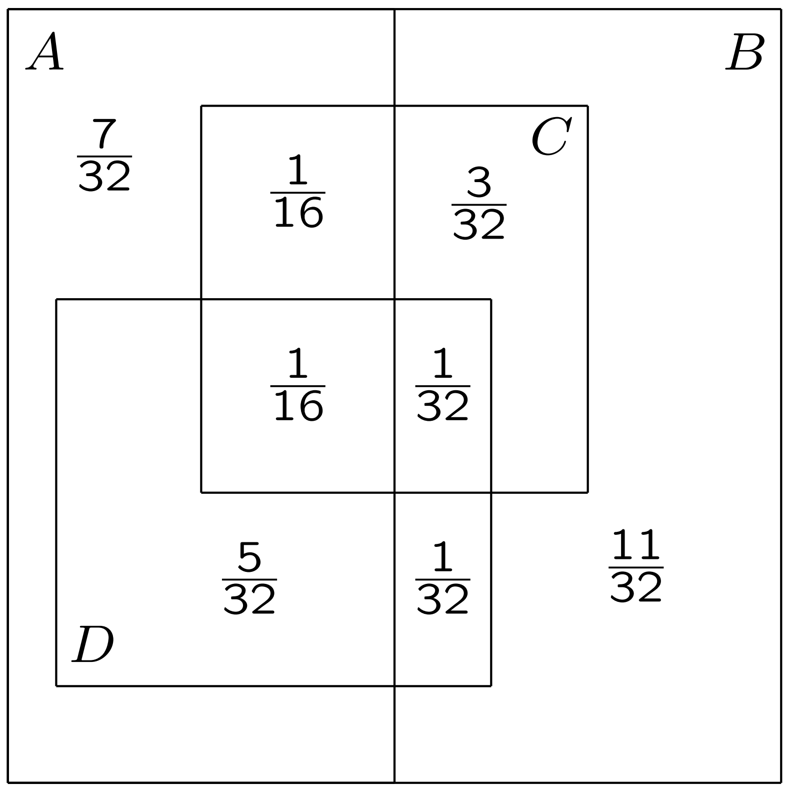


- A and C are independent because

$$\begin{aligned}\Pr[A \cap C] &= \frac{1}{2} \times \frac{1}{4} \\ &= \Pr[A]\Pr[C].\end{aligned}$$

- Are B and C independent $\boxed{\quad ?}$

More Independent Events?



- Are A and D independent $\boxed{\quad ?}$
- Are C and D independent $\boxed{\quad ?}$

Partition of Sample Space

A set of events $\{A_1, A_2, \ldots, A_n\}$ is a **partition** of Ω if

1. Whenever $i \neq j$, then

$$A_i \cap A_j = \emptyset;$$

2. For all i, where $1 \leq i \leq n$,

$$\Pr[A_i] > 0;$$

3.

$$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega.$$

Corresponds to **case analysis.**

Law of Total Probability

Suppose $\{A_1, A_2, \ldots, A_n\}$ is a partition of Ω. Then, for any event $A \subseteq \Omega$, we have

$$\Pr[A] \;=\; \sum_{i=1}^{n} \Pr[A_i] \Pr[A \mid A_i] \,.$$

PROOF.

Combine

$$\Pr[A] \;=\; \sum_{i=1}^{n} \Pr[A \cap A_i]$$

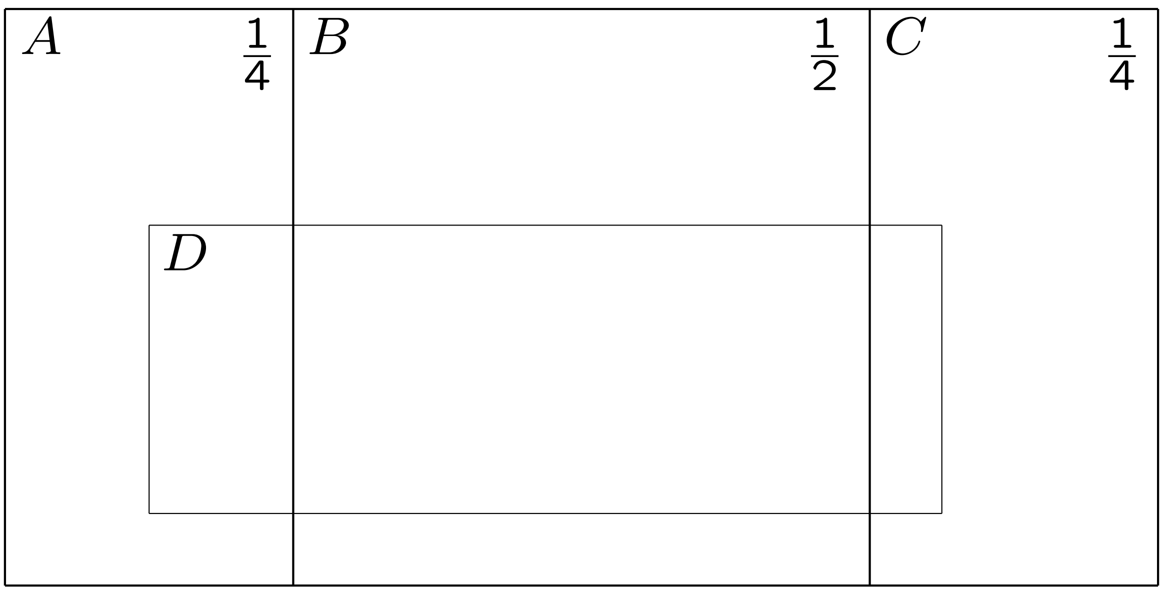
with

$$\Pr[A \cap A_i] \;=\; \Pr[A_i] \Pr[A \mid A_i] \,.$$

$\square$

Example

Partition $\{A, B, C\}$ of Ω:



By Law of Total Probability,

$$\begin{aligned}
\Pr[D] &= \Pr[A]\Pr[D \mid A] + \Pr[B]\Pr[D \mid B] \\
&\quad + \Pr[C]\Pr[D \mid C] \\
&= \frac{1}{4} \times \frac{1}{4} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{8} \\
&= \frac{11}{32}.
\end{aligned}$$

Bayes Theorem

Idea is to reverse the conditions in the conditional probabilities.

$$\begin{aligned} \Pr[B \mid A] &= \frac{\Pr[B \cap A]}{\Pr[A]} \\ &= \frac{\Pr[B]\Pr[A \mid B]}{\Pr[A]} \end{aligned}$$

Use to solve for fourth probability when three are known.

Trivedi Example 1.12

At the Triangle Universities Computation Center, 15% of the jobs are from Duke, 35% from UNC, and 50% from NCSU. Suppose the probabilities that a job requires operator setup are 0.01, 0.05, and 0.02, respectively.

1. What is the probability that a random job requires operator setup?

2. What is the probability that a random job requiring operator setup came from UNC?

Events: A_{setup}, B_{Duke}, B_{UNC}, B_{NCSU}.

1. By Law of Total Probability,

$$\begin{aligned}
&\Pr\left[A_{\text{setup}}\right] \\
&\quad = \Pr\left[B_{\text{Duke}}\right]\Pr\left[A_{\text{setup}} \mid B_{\text{Duke}}\right] \\
&\qquad + \Pr\left[B_{\text{UNC}}\right]\Pr\left[A_{\text{setup}} \mid B_{\text{UNC}}\right] \\
&\qquad + \Pr\left[B_{\text{NCSU}}\right]\Pr\left[A_{\text{setup}} \mid B_{\text{NCSU}}\right] \\
&\quad = (0.15)(0.01) + (0.35)(0.05) \\
&\qquad + (0.50)(0.02) \\
&\quad = 0.029.
\end{aligned}$$

2 By Bayes Theorem,

$$
\begin{aligned}
&\Pr\left[B_{\text{UNC}} \mid A_{\text{setup}}\right] \\
&\quad = \frac{\Pr\left[B_{\text{UNC}}\right]\Pr\left[A_{\text{setup}} \mid B_{\text{UNC}}\right]}{\Pr\left[A_{\text{setup}}\right]} \\
&\quad = \frac{(0.35)(0.05)}{0.029} \\
&\quad \approx 0.603.
\end{aligned}
$$

Random Variables

Event space Ω

Probability measure Pr

Random variable X

- Random variable X is a real function defined on the elementary events.
- Corresponds to a measurement of a random model or sample.
- Probabilities $\Pr[X = 5]$ and $\Pr[-7.5 \leq X < 8.75]$ are well-defined.

Counting Example

Sample space consists of variable-length messages γ (bit-strings). Think of a message as received one bit at a time. The probability that the next bit received is the last is uniformly ρ ($0 < \rho < 1$). Let the random variable X be the length of the message:

$$X(\gamma) \;=\; |\gamma|.$$

- X is a **discrete** random variable.

- We compute

$$\Pr[X = n] \;=\; (1 - \rho)^{n-1}\rho.$$

- $X(\gamma) = n$ holds for exactly 2^n elementary events (messages).

Continuous Measurement

Communication time between two nodes in the campus network is random. Let X be the random variable that gives the time for a certain message to transit from `ap1.cs.vt.edu` to `arabidopsis.cs.vt.edu`.

- $\Pr[X = t] = 0$ for any t.
- X is a **continuous** random variable. It can take on an uncountable number of values.
- Probability on intervals makes more sense: $\Pr[a \leq X \leq b]$ may well be positive.

Computing Probabilities

Intuitively

$$\Pr[a \leq X \leq b] = \sum_{\substack{x \\ a \leq x \leq b}} \Pr[x].$$

But this only works for discrete random variables. We use some auxiliary functions to help us compute probabilities.

- **Probability Mass Function**

$$p_X(x) = \Pr[X = x]$$

- **(Cumulative) Distribution Function**

$$F_X(x) = \Pr[X \leq x]$$

- **Density Function**

$$f_X(x) = \frac{dF_X(x)}{dx},$$

where defined.

Auxiliary Functions

Event space Ω

Probability measure Pr

Random variable X

Probability Mass Function (Discrete only)

Distribution Function (Discrete and Continuous)

Density Function (Continuous only)

Example One

Two branch banks, b_1 and b_2, submit transactions to central accounting daily. The number of transactions submitted by either has a uniform distribution between 1 and 1000 daily. Let X_1 and X_2 be the number of transactions from b_1 and b_2, respectively, today.

- Discrete or continuous $\boxed{\quad ?}$
- $p_{X_1}(n) = p_{X_2}(n) = \boxed{\quad ?}$
- $$F_{X_1}(n) = \begin{cases} 0 & n \leq 0 \\ \dfrac{n}{1000} & 1 \leq n \leq 1000 \\ 1 & 1000 < n \end{cases}$$

Example One (Continued)

- Let $X = X_1 + X_2$, the total number of transactions received today.

$$
\begin{aligned}
p_X(n) & \\
&= \boxed{\quad ?} \\
&= \begin{cases} 0 & n < 2 \text{ or } n > 2000 \\ \dfrac{n-1}{(1000)^2} & 2 \le n \le 1001 \\ \dfrac{2001-n}{(1000)^2} & 1002 \le n \le 2000 \end{cases}
\end{aligned}
$$

- Can you plot $p_X(n)$ and $F_X(n)$?

Example Two

Suppose a component C fails at a time given by an exponential random variable with parameter α. Let X be the random variable that equals the **time of failure.**

Is X discrete or continuous $\boxed{\quad ?}$

The distribution function for X is

$$F_X(t) = \Pr[C \text{ fails by time } t] = \begin{cases} 0 & t \le 0 \\ 1 - e^{-\alpha t} & 0 < t \end{cases}$$

Verify that this is a distribution function.

Example Two (Continued)

- The density function for X is

$$
\begin{aligned}
f_X(t) &= \frac{dF_X(t)}{dt} \\
&= \begin{cases} 0 & t \leq 0 \\ \alpha e^{-\alpha t} & 0 < t \end{cases}
\end{aligned}
$$

- To compute the probability that C fails in an interval $[a, b]$, where $0 \leq a \leq b$, just integrate the density function

$$
\begin{aligned}
\Pr[a \leq X \leq b] &= \int_a^b f_X(t)\, dt \\
&= e^{-\alpha a} - e^{-\alpha b}
\end{aligned}
$$

- Verify that

$$\int_{-\infty}^{\infty} f_X(t)\, dt = 1.$$

Note on the Exponential Distribution

- Note that the density function for exponential R.V. X with parameter α can be defined with either $f_X(0) = 0$ or $f_X(0) = \alpha$. Actually ANY finite value will do. A difference in value between two density functions at only a countable number of points never makes a difference in the distribution function.

Example Two (Continued)

Conditional Distribution Function

Suppose $0 \leq R \leq t$. If we know that C is still functioning at time R, we may be interested in the conditional distribution function for X.

Conditional distribution function:

$$
\begin{aligned}
&\Pr[X \leq t \mid R \leq X] \\
&\quad = \frac{\Pr[X \leq t \text{ and } R \leq X]}{\Pr[R \leq X]} \\
&\quad = \frac{\Pr[R \leq X \leq t]}{\Pr[R \leq X]}
\end{aligned}
$$

Example Two (Concluded)

- **Conditional distribution function:**

$$
\begin{aligned}
\Pr[X \le t \mid R \le X] &= \frac{F_X(t) - F_X(R)}{1 - F_X(R)} \\
&= \frac{e^{-\alpha R} - e^{-\alpha t}}{e^{-\alpha R}} \\
&= 1 - e^{-\alpha(t-R)}
\end{aligned}
$$

- We just get the unconditional distribution function for X shifted by R:

$$
\Pr[X \le t \mid R \le X] = F_X(t - R).
$$

  The exponential distribution is called **memoryless** for this reason.

Exercise

Let X be a random variable with a uniform continuous distribution on the interval $[a, b]$, where $a < b$.

- What is the probability mass function for X $\boxed{\quad ?}$

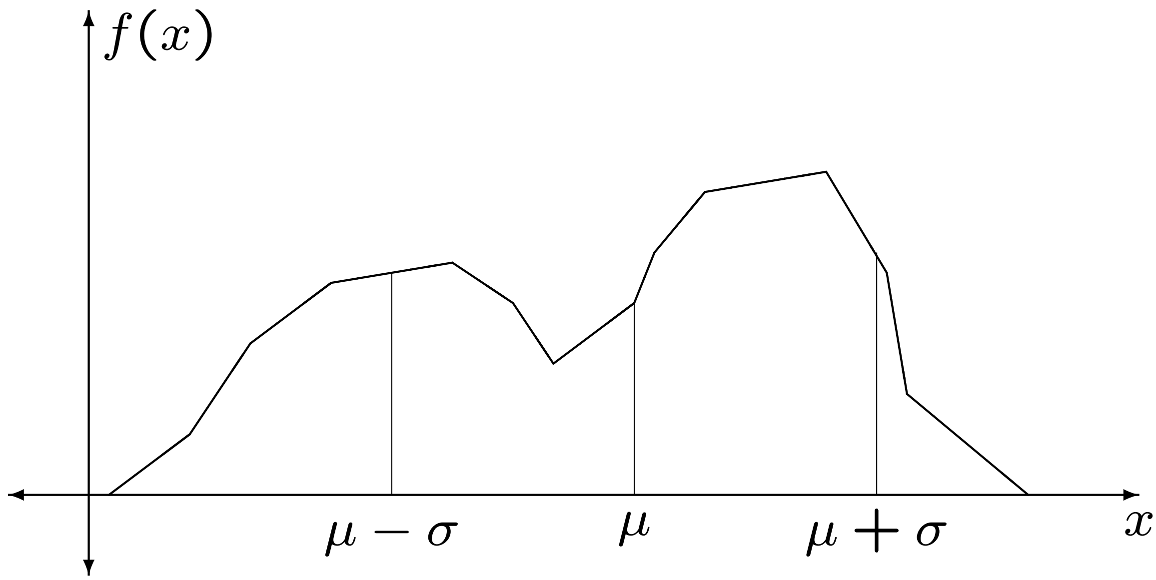
- What is the density function for X $\boxed{\quad ?}$

- What is the distribution function for X $\boxed{\quad ?}$

- Suppose $a < c < (a + b)/2$. Compute

$$\Pr[X \leq c \mid X \leq (a+b)/2] \quad = \quad \boxed{\quad ?}$$

Expected Value and Variance



- **Expected Value** μ measures the "central value" of the distribution. Also called **mean value** or **average value.**

- **Variance** σ^2 measures the "spread" of the distribution.

Expected Value

- **Discrete Random Variable X.** Let $x_1, x_2, x_3, \ldots$ be the finite or infinite sequence of points x at which $p_X(x) > 0$.

$$\mathsf{E}\,[X] \quad = \quad \sum_{x_i} x_i p_X(x_i),$$

provided

$$\sum_{x_i} |x_i| p_X(x_i) \quad < \quad \infty.$$

- **Continuous Random Variable Y.**

$$\mathsf{E}\,[Y] \quad = \quad \int_{-\infty}^{\infty} y f_Y(y)\, dy,$$

provided

$$\int_{-\infty}^{\infty} |y| f_Y(y)\, dy \quad < \infty.$$

- Alternate notation for expected value is $\mu = \mu_X$.

Variance

- **Variance** for random variable X:

$$\mathsf{Var}[X] \quad = \quad \mathsf{E}\left[(X - \mu_X)^2\right],$$

  if the expected value is defined.

- Alternate notation for variance is $\sigma^2 = \sigma_X^2$.

- Square root of variance is σ, the **Standard Deviation.**

Variance Particulars

- **Discrete Random Variable X.** Let $x_1, x_2, x_3, \ldots$ be the finite or infinite sequence of points x at which $p_X(x) > 0$.

$$\mathsf{Var}[X] \quad = \quad \sum_{x_i} (x_i - \mathsf{E}[X])^2 p_X(x_i),$$

  provided the sum is finite.

- **Continuous Random Variable Y.**

$$\mathsf{Var}[Y] \quad = \quad \int_{-\infty}^{\infty} (y - \mathsf{E}[Y])^2 f_Y(y)\, dy,$$

  provided the integral is finite.

Uniform Discrete Random Variable

Discrete random variable X has positive probability at n points $x_1, x_2, \ldots, x_n$. Then

$$p_X(x) = \begin{cases} \frac{1}{n} & x \in \{x_1, x_2, \ldots, x_n\} \\ 0 & \text{otherwise} \end{cases}$$

- **Expected Value.**

$$\begin{aligned} \mathsf{E}[X] &= \sum_{i=1}^{n} x_i p_X(x_i) \\ &= \frac{\sum_{i=1}^{n} x_i}{n}. \end{aligned}$$

Uniform Discrete Random Variable Continued

Suppose $x_i = i$.

- **Expected Value.**

$$\begin{aligned} \mathsf{E}[X] &= \frac{\sum_{i=1}^{n} x_i}{n} \\ &= \frac{n(n+1)/2}{n} \\ &= \frac{n+1}{2} \end{aligned}$$

Uniform Discrete Random Variable Concluded

- **Variance.**

$$\begin{aligned}
\text{Var}[X] &= \sum_{i=1}^{n} \frac{(i-(n+1)/2)^2}{n} \\
&= \frac{(n-1)(n+1)}{12}
\end{aligned}$$

- **Standard Deviation.** For large n,

$$\begin{aligned}
\sigma_X &\approx \frac{n}{\sqrt{12}} \\
&\approx \frac{n}{3.46}.
\end{aligned}$$

Uniform Continuous Random Variable

Y has a uniform continuous distribution over the interval $[a, b]$.

$$f_Y(y) = \begin{cases} \dfrac{1}{b-a} & a \leq y \leq b \\ 0 & \text{otherwise} \end{cases}$$

Expected Value.

$$\begin{aligned} \mathsf{E}[Y] &= \int_{-\infty}^{\infty} y f_Y(y)\, dy \\ &= \int_a^b \frac{y}{b-a}\, dy \\ &= \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$

Uniform Continuous Random Variable Continued

Variance.

$$
\begin{aligned}
\mathsf{Var}[Y] &= \int_{-\infty}^{\infty} (y - \mathsf{E}[Y])^2 f_Y(y)\, dy \\
&= \int_a^b \frac{(y - \mathsf{E}[Y])^2}{b-a}\, dy \\
&= \frac{1}{b-a} \int_a^b \left(y - \frac{b+a}{2} \right)^2 dy \\
&= \frac{1}{b-a} \int_a^b y^2 - 2y\frac{b+a}{2} + \left(\frac{b+a}{2} \right)^2 dy \\
&= \frac{1}{b-a} \left[\frac{b^3 - a^3}{3} - \frac{(b^2 - a^2)(b+a)}{2} \right. \\
&\qquad \left. + \frac{(b-a)(b+a)^2}{4} \right]
\end{aligned}
$$

Uniform Continuous Random Variable Concluded

$$
\begin{aligned}
\text{Var}\,[Y] & \\
= \quad & \frac{1}{b-a}\left[\frac{b^3-a^3}{3} - \frac{(b^2-a^2)(b+a)}{2}\right. \\
& \left. + \frac{(b-a)(b+a)^2}{4}\right] \\
= \quad & \frac{b^2+ab+b^2}{3} - \frac{b^2+2ab+a^2}{4} \\
= \quad & \frac{4b^2+4ab+4a^2-(3b^2+6ab+3a^2)}{12} \\
= \quad & \frac{(b-a)^2}{12}
\end{aligned}
$$

Standard Deviation

$$
\sigma \quad = \quad \frac{b-a}{\sqrt{12}}
$$

Exponential Random Variable

Y has density function

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \alpha e^{-\alpha y} & 0 \leq y \end{cases}$$

- **Expected Value.**

$$\mathsf{E}[Y] = \int_0^\infty y \alpha e^{-\alpha y} \, dy$$

Need to evaluate the indefinite integral

$$\begin{aligned} \int y \alpha e^{-\alpha y} \, dy &= \int -y \, d\left(e^{-\alpha y}\right) \\ &= -y e^{-\alpha y} - \int e^{-\alpha y} \, d(-y) \\ &= -y e^{-\alpha y} + \int e^{-\alpha y} \, dy \\ &= -y e^{-\alpha y} - \frac{e^{-\alpha y}}{\alpha} \end{aligned}$$

Exponential Random Variable Continued

- Back to expected value:

$$\begin{aligned}
\mathsf{E}[Y] &= \int_0^{\infty} y\alpha e^{-\alpha y}\,dy \\
&= -(\infty)e^{-\alpha(\infty)} - \frac{e^{-\alpha(\infty)}}{\alpha} \\
&\quad - \left(-0e^{-\alpha 0} - \frac{e^{-\alpha 0}}{\alpha}\right) \\
&= \frac{1}{\alpha}
\end{aligned}$$

- Note the use of L'Hospital's Rule in evaluating

$$(\infty)e^{-\alpha(\infty)}.$$

Exponential Random Variable Concluded

- **Variance.** A similar calculation gives

$$\begin{aligned}\text{Var}[Y] &= \int_0^\infty \left(y - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha y}\, dy \\ &= \frac{1}{\alpha^2}\end{aligned}$$

- *You should be able to do this calculation.*

Properties of Expectation

Theorem. Suppose X and Y are random variables such that $\mathsf{E}[X]$ and $\mathsf{E}[Y]$ are defined.

1. For all real β, $\mathsf{E}[\beta X] = \beta \mathsf{E}[X]$.

2. $\mathsf{E}[X+Y] = \mathsf{E}[X] + \mathsf{E}[Y]$.

3. If X and Y are independent, then $\mathsf{E}[XY] = \mathsf{E}[X]\,\mathsf{E}[Y]$.

For the proof, see Theorem 2.7.1.

Properties 1 and 2 together say that **expectation is linear.**

Application of Linearity

Recall from Page 33 the two branch banks that submitted X_1 and X_2 transactions daily.

$p_{X_1}(n) = p_{X_2}(n) = 1/1000$, where $1 \le n \le 1000$.

$$p_{X_1+X_2}(n) = \begin{cases} 0 & n < 2 \text{ or } n > 2000 \\ \dfrac{n-1}{(1000)^2} & 2 \le n \le 1001 \\ \dfrac{2001-n}{(1000)^2} & 1002 \le n \le 2000 \end{cases}$$

$\mathrm{E}[X_1] = 1001/2$

$\mathrm{E}[X_1 + X_2] = \boxed{\quad ?}$

Another Application

Find a higher-level expression for variance using linearity of expectation.

$$
\begin{aligned}
&\mathrm{Var}[X] \\
&= \quad \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\
&= \quad \mathrm{E}\left[X^2 - 2X\mathrm{E}[X] + \mathrm{E}[X]^2\right] \\
&= \quad \mathrm{E}\left[X^2\right] - 2\mathrm{E}[X]\,\mathrm{E}[X] + \mathrm{E}[X]^2 \\
&= \quad \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 \\
&= \quad \mathrm{E}\left[X^2\right] - \mu^2
\end{aligned}
$$

This is often easier to apply than the "defining" formula for variance.

Discrete Random Variables

- **Discrete Probability Distributions.** Random variable assumes only a countable number of values.

- Distribution function of a R.V. summarizes the underlying sample space as far as the R.V. is concerned. Result is to ignore the complexities of the sample space.

- We consider only a few specific distributions.

Basic Discrete Random Variables

- **Discrete Uniform R.V.** Assumes only a finite number of values, each with the same probability.

- **Bernoulli R.V.** There are only two values:
  - **Success (1)** with probability p.
  - **Failure (0)** with probability $q = 1 - p$.

Other Discrete Random Variables

- **Binomial R.V.** Based on a finite sequence of independent Bernoulli trials. Counts successes.

- **Geometric R.V.** Based on an infinite sequence of independent Bernoulli trials. Detects first success.

- **Poisson R.V.** Not Bernoulli based. Takes on an infinite set of values.

Binomial Random Variable

- Binomial R.V. X counts the number of successes among n independent Bernoulli trials. Parameterized by n and p.
- Assumes values $0, 1, 2, \ldots, n$.
- **Probability Mass Function.** For $0 \leq k \leq n$,

$$\begin{aligned} \Pr[X = k] &= b(k; n, p) \\ &= \binom{n}{k} p^k q^{n-k} \end{aligned}$$

Parameters of Binomial R.V.

- **Sum of Bernoulli R.V.'s.**
  $X = X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent, identically distributed Bernoulli R.V.'s.

- **Expected Value.**

$$\begin{aligned} \mathsf{E}[X] &= \sum_{i=1}^{n} \mathsf{E}[X_i] \\ &= np \end{aligned}$$

- **Variance.**

$$\mathsf{Var}[X] = npq$$

Shape of the Binomial R.V.

For $1 \leq k \leq n$, consider the ratio r_k of consecutive probabilities:

$$
\begin{aligned}
r_k &= \frac{b\,(k; n, p)}{b\,(k-1; n, p)} \\
&= \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k-1} p^{k-1} q^{n-k+1}} \\
&= \frac{\dfrac{n!}{k!(n-k)!}}{\dfrac{n!}{(k-1)!(n-k+1)!}} \left(\frac{p}{q}\right) \\
&= \left(\frac{n-k+1}{k}\right) \left(\frac{p}{q}\right)
\end{aligned}
$$

Shape Continued

Consider the sequence $r_1, r_2, \ldots, r_n$:

$$
\begin{aligned}
r_k &= \left(\frac{n-k+1}{k}\right)\left(\frac{p}{q}\right) \\
r_1 &= n(p/q) \\
r_n &= (1/n)(p/q)
\end{aligned}
$$

There are three cases:

Condition	**Shape of** $b\,(k; n, p)$
$n \leq q/p$	Max at $k = 0$; decreases monotonically.
$r_t \geq 1 > r_{t+1}$	Max at $k = t$; increases monotonically to t, then decreases monotonically.
$n \leq p/q$	Max at $k = n$; increases monotonically.

Geometric R.V.

In an infinite sequence of Bernoulli trials, let X be the number of trials **before** the first success.

- X assumes values $0, 1, 2, \ldots$.

- **Probability Mass Function.**

$$\Pr[X = k] \quad = \quad q^k p$$

Geometric R.V. Continued

- **Expected Value.**

$$\mathrm{E}[X] = \frac{q}{p}$$

- **Variance.**

$$\mathrm{Var}[X] = \frac{q}{p^2}$$

- **Memoryless,** like an exponential R.V.

Poisson R.V.

If X has a Poisson distribution with parameter α, then its probability mass function is

$$\Pr[X = k] \quad = \quad e^{-\alpha}\frac{\alpha^k}{k!}.$$

- X assumes values $0, 1, 2, \ldots$.

Poisson R.V. Continued

- **Expected Value.**

$$
\begin{aligned}
\mathsf{E}[X] &= \psi'_X[0] \\
&= \alpha e^{\theta} e^{\alpha(e^{\theta}-1)} \Big|_{\theta=0} \\
&= \alpha
\end{aligned}
$$

- **Variance.**

$$
\begin{aligned}
\mathsf{E}\left[X^2\right] &= \psi''_X[0] \\
&= \alpha(1+\alpha e^{\theta}) e^{\theta+\alpha(e^{\theta}-1)} \Big|_{\theta=0} \\
&= \alpha(1+\alpha) \\
\mathsf{Var}[X] &= \mathsf{E}\left[X^2\right] - (\mathsf{E}[X])^2 \\
&= \alpha
\end{aligned}
$$

- **Applications:** later.

Continuous Random Variables

- **Continuous Probability Distributions.** Random variable assumes an uncountable number of values.

- For our purposes, computing the probability that a continuous R.V. takes on a range of values corresponds to **integrating the density function** over that range of values.

- Typically in computer science applications, the value of a continuous R.V. corresponds to **time** in the future, with the current time being time 0.

- Less frequently, the value is **spatial** (length, area, etc.).

Continuous Random Variables

- **Continuous Uniform R.V.** Assumes values only in a bounded interval, with constant density function for that interval.

- **Exponential R.V.** Exponentially decreasing density function.

Uniform Continuous Random Variable

Y has a uniform continuous distribution over the interval $[a, b]$.

$$f_Y(y) = \begin{cases} \dfrac{1}{b-a} & a \le y \le b \\ 0 & \text{otherwise} \end{cases}$$

Expected Value.

$$\begin{aligned} \mathsf{E}[Y] &= \int_a^b \frac{y}{b-a}\, dy \\ &= \frac{b+a}{2} \end{aligned}$$

Variance.

$$\begin{aligned} \mathrm{Var}[Y] &= \int_a^b \frac{(y - \mathsf{E}[Y])^2}{b-a}\, dy \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Exponential R.V.

Y has density function

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \alpha e^{-\alpha y} & 0 \leq y \end{cases}$$

Expected Value.

$$\begin{aligned} \mathsf{E}[Y] &= \int_0^\infty y \alpha e^{-\alpha y}\, dy \\ &= \frac{1}{\alpha} \end{aligned}$$

Variance.

$$\begin{aligned} \mathsf{Var}[Y] &= \int_0^\infty \left(y - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha y}\, dy \\ &= \frac{1}{\alpha^2} \end{aligned}$$

Memoryless.

$$\Pr[Y \leq t \mid R \leq Y] = F_Y(t - R).$$

Modeling Constant Rate of Point Mutations

- Use exponential distribution, since it is memoryless.
- Parameter α is the rate at which mutations occur.
- $\mathrm{E}[Y] = 1/\alpha$ makes sense.
- Once one mutation occurs, the process starts all over again.
- In an interval of length β, the number of mutations follows a Poisson distribution with parameter β/α.

Stochastic Processes

- A **stochastic process** $\mathsf{X} = \{X(t) : t \in T\}$ is a family of random variables indexed by some set T. Often T is time.

- The set of all possible values of the $X(t)$ is the **state space** of X.

- If T is uncountable (e.g., $T = R^+ = \{t : t \geq 0\}$), then X is a **continuous parameter** process.

- If T is countable (e.g., $T = Z$ or $T = Z^+ = \{n : n \in Z \text{ and } n \geq 0\}$), then X is a **discrete parameter** process.

Stochastic Processes from Message Arrivals

Process	Index Set	State Space
$\mathrm{I} = \{\tau_n\}$	$\{1, 2, 3, \ldots\}$ Discrete	R^+ Continuous
$\mathrm{R} = \{\lambda_{a,b}\}$	T_{R} Continuous	Z^+ Discrete
$\mathrm{M} = \{M_n\}$	Z^+ Discrete	R^+ Continuous
$\mathrm{N} = \{N(t)\}$	R^+ Continuous	Z^+ Discrete

$$T_{\mathrm{R}} = \{(a, b) : 0 \leq a < b\}$$

Counting Process

A stochastic process $\mathsf{X} = \{X(t) : t \in R^+\}$ is a **counting process** for a set of events occurring in the time interval $(0, \infty)$ if

- $X(0) = 0$.
- $X(t_2) - X(t_1)$ is the number of events occurring in the interval $(t_1, t_2]$.

As a consequence, we have

- $X(t)$ assumes values in Z^+.
- $t_1 < t_2$ implies that $X(t_1) \leq X(t_2)$.

A Counting Process

The message counting process
N $= \{N(t) : t \in R^+\}$ is a counting process.

We want to investigate the kinds of properties N might (or should) have.

- Should $N(t_1)$ and $N(t_2)$, where $t_1 < t_2$, be independent?

- What does it mean for messages to arrive at a **constant rate?** Two possibilities:

  1. Constantly spaced messages. E.g., message m_n arrives **precisely** at time n seconds.

  2. The probability of one event in a small interval is constant everywhere. Let's develop some terminology so we can talk about this possibility.

Definitions

- The function $f(h)$ is $o(h)$ if

$$\lim_{h \to 0^+} \frac{f(h)}{h} = 0.$$

- $\mathsf{N} = \{N(t) : t \in R^+\}$ has **independent increments** if whenever $(a_1, b_1]$ and $(a_2, b_2]$ are nonoverlapping intervals, then $N(b_1) - N(a_1)$ and $N(b_2) - N(a_2)$ are independent R.V.s.

- $\mathsf{N} = \{N(t) : t \in R^+\}$ has **stationary increments** if for all $t_1 \geq 0$, $t_2 \geq 0$, and $h \geq 0$, the random variables

$$N(t_1 + h) - N(t_1)$$

and

$$N(t_2 + h) - N(t_2)$$

have identical distribution functions.

Notes for the Definitions

Caution. The little-oh defined here is not the little-oh of algorithm analysis. In particular, algorithm analysis concentrates on $h \to \infty$.

A function that is $o(h)$ should be thought of as small in the upcoming definition. The definition is a formal statement of that intuitive notion.

Poisson Process

Definition 4.2.1. A counting process N $= \{N(t) : t \in R^{+}\}$ is a **Poisson process with rate** $\lambda > 0$ if these properties hold.

1. N has independent increments.

2. N has stationary increments.

3. The probability of one event occurring in a small interval of length h is approximately λh:

$$\Pr[N(h) = 1] \quad = \quad \lambda h + o(h).$$

4. The probability of two or more events occurring in a small interval of length h is approximately 0:

$$\Pr[N(h) \geq 2] \quad = \quad o(h).$$

Poisson Process

Theorem 4.2.1. Let N $= \{N(t) : t \in R^+\}$ be a Poisson process with rate $\lambda > 0$. Fix $t > 0$. Let the random variable Y be the number of events within an interval of length t. Then Y has a Poisson distribution with parameter λt. That is,

$$\Pr[Y = k] \quad = \quad e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

for $k \in Z^+$.

In particular, the expected number of events in the interval $(0, t]$ is $\mathrm{E}[Y] = \lambda t$, exactly what one would expect of a **constant rate counting process.**

Interarrival Times

Theorem 4.2.2. Let N $= \{N(t) : t \in R^+\}$ be a Poisson process with rate $\lambda > 0$. Let τ_n be the nth interarrival time, as before. Then the τ_n's are independent, identically distributed exponential R.V.s with parameter λ.

PROOF. Since N has independent increments, the τ_n's are independent.

Since N has stationary increments, we may calculate

$$\begin{aligned} \Pr[\tau_n > s] &= \Pr[N(s) = 0] \\ &= e^{-\lambda s}. \end{aligned}$$

Hence

$$\Pr[\tau_n \leq s] = 1 - e^{-\lambda s},$$

so τ_n is an exponential R.V. with parameter λ. $\square$

The Converse

Theorem 4.2.3. Let N $= \{N(t) : t \in R^+\}$ be a counting process such that the τ_n's are independent, identically distributed exponential R.V.s with parameter λ. Then N is a Poisson process with rate λ.

PROOF. Since the τ_n's are memoryless, we get independent increments and stationary increments.

As before, let M_n be the time of the nth event. By Theorem 3.2.6, M_n has a gamma distribution with parameters n and λ. Hence

$$
\begin{aligned}
\Pr[N(t) = n] &= \Pr[M_n \le t] - \Pr[M_{n+1} \le t] \\
&= \int_0^t \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}\, dx \\
&\quad - \int_0^t \frac{\lambda(\lambda x)^{n} e^{-\lambda x}}{n!}\, dx
\end{aligned}
$$

Proof Continued

$$
\begin{aligned}
\Pr\left[N(t) = n\right] &= \int_0^t \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}\,dx \\
&\quad - \int_0^t \frac{\lambda(\lambda x)^{n} e^{-\lambda x}}{n!}\,dx \\
&= \int_0^t \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}\,dx \\
&\quad - \int_0^t \frac{-(\lambda x)^{n}}{n!}\,d\left(e^{-\lambda x}\right) \\
&= \int_0^t \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}\,dx \\
&\quad + \frac{(\lambda x)^{n} e^{-\lambda x}}{n!}\Bigg|_0^t \\
&\quad - \int_0^t \frac{e^{-\lambda x}}{n!}\,d\left((\lambda x)^{n}\right)
\end{aligned}
$$

Proof Concluded

$$
\begin{aligned}
&\Pr[N(t) = n] \\
&= \int_0^t \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}\, dx + \frac{(\lambda x)^n e^{-\lambda x}}{n!} \Bigg|_0^t \\
&\quad - \int_0^t \frac{e^{-\lambda x}}{n!}\, d\left((\lambda x)^n\right) \\
&= \int_0^t \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}\, dx + \frac{(\lambda t)^n}{n!} e^{-\lambda t} \\
&\quad - \int_0^t \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}\, dx \\
&= \frac{(\lambda t)^n}{n!} e^{-\lambda t}
\end{aligned}
$$

Hence $N(t)$ has a Poisson distribution with parameter λt. We conclude that N is a Poisson process.

$\square$

Constant Rate Arrival = Poisson Process

We should remember a Poisson process as the counting process that corresponds to the arrival of messages (or customers) at a server with a **constant rate**.

This is the most tractable assumption for modeling message arrival.

Graphical Representation

Poisson Process. An infinite graph (actually a path) with each state of the counting process a node and the arcs labeled with the arrival rate.

$$0 \xrightarrow{\lambda} 1 \xrightarrow{\lambda} 2 \xrightarrow{\lambda} 3 \xrightarrow{\lambda} 4 \xrightarrow{\lambda} \quad \cdots$$

More General Counting Process. Each arc could have a different arrival rate.

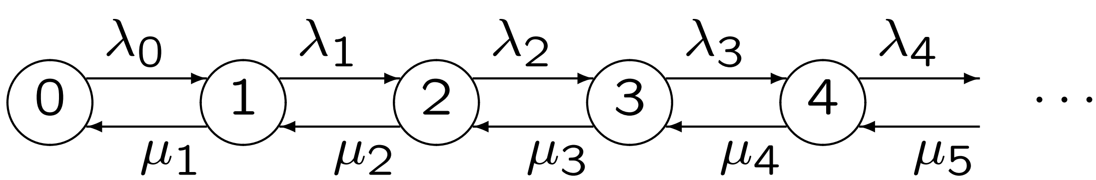
$$0 \xrightarrow{\lambda_0} 1 \xrightarrow{\lambda_1} 2 \xrightarrow{\lambda_2} 3 \xrightarrow{\lambda_3} 4 \xrightarrow{\lambda_4} \quad \cdots$$

Queuing

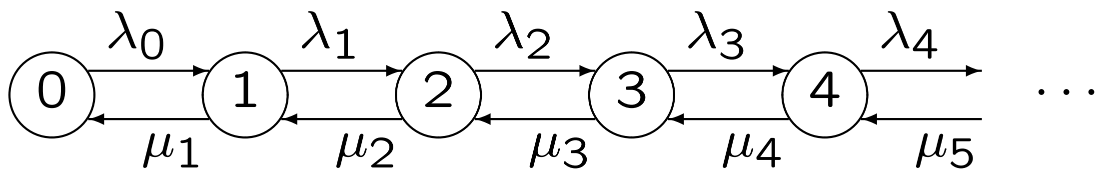
- A communication switch typically has a queue to store incoming messages **(arrivals or births)** until they can be retransmitted **(departures or deaths).**
- There can be a **rate of departure,** just as there is a rate of arrival.
- No longer a counting process, but a slightly more general **birth-and-death process** $\mathsf{X} = \{X(t) : t \geq 0\}$.
- Still a continuous parameter, discrete state space, stochastic process.
- A **birth** takes the process from state n to state $n + 1$, while a **death** takes the process from state n to state $n - 1$. A death cannot occur in state 0.

Representation

Assume a birth occurs from state n to $n+1$ at the rate λ_n, while a death occurs from state n to $n-1$ at the rate μ_n. There are now two transitions (arcs) out of every state except 0. Get the **state-transition rate diagram.**



Interpretation



- State n represents the accumulation of n messages in the communications switch.

- Typically assume process starts in state 0.

- If there is any state n for which a birth in state n is disallowed, then the process has a finite number of states (at most $n+1$).

- Some deaths may be disallowed also. Example, a Poisson process!

- Note the resemblance to a (finite) automata or state machine.

Analysis of Birth and Death Processes

Summary. We have a stochastic process $\mathsf{X} = \{X(t) : t \geq 0\}$ with state space Z^{+}, birth rates λ_n, $n \geq 0$, and death rates μ_n, $n \geq 1$. Assume initially in state 0.

- $\Pr[X(t) = n]$ is the probability that at time t the process is in state n. Under some circumstances (e.g., a Poisson process), the distribution of $X(t)$ can be analytically derived.

- **Limiting Probabilities.** For all $n \in Z^{+}$, define

$$p_n = \lim_{t\to\infty} \Pr[X(t) = n],$$

if the limit exists.

Stationary Process

If all the limits

$$p_n \quad = \quad \lim_{t \to \infty} \Pr\left[X(t) = n\right]$$

exist, then X is a **stationary process** and reaches a **steady state** distribution given by the p_n's (in the limit).

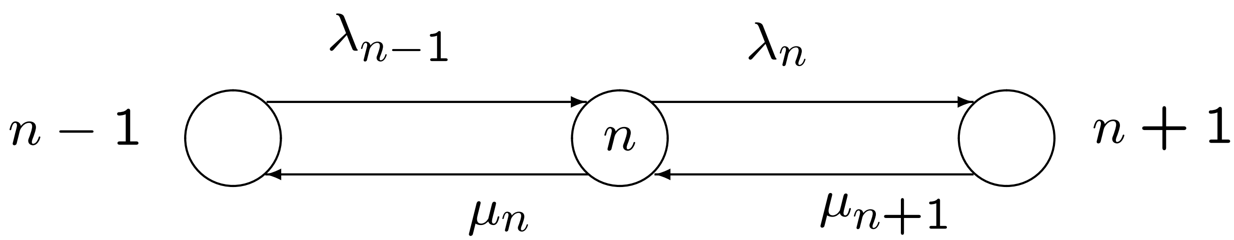
- If every $p_n = 0$, then X moves off to an "infinite state," a disaster if X represents the messages contained in a communications switch!

- Otherwise,

$$\sum_{n=0}^{\infty} p_n \quad = \quad 1.$$

- How can we determine the p_n?

Flow



- **Flow In.**

  FLOWIN_n

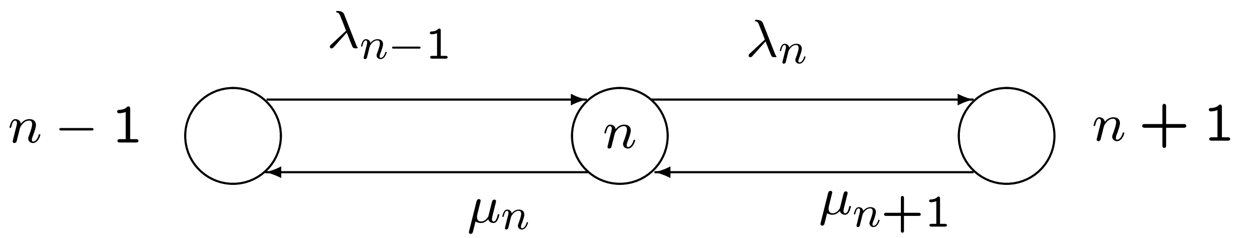
$$= \begin{cases} p_{n-1}\lambda_{n-1} + p_{n+1}\mu_{n+1} & n > 0 \\ p_{n+1}\mu_{n+1} & n = 0 \end{cases}$$

- **Flow Out.**

  FLOWOUT_n

$$= \begin{cases} p_n(\lambda_n + \mu_n) & n > 0 \\ p_n\lambda_n & n = 0 \end{cases}$$

Balance Equations



In the limit, the rate of flow (births and deaths) into state n must equal the rate of flow out.

- The **balance equations** are, for all $n \geq 0$,

$$\mathrm{FLOWIN}_n = \mathrm{FLOWOUT}_n.$$

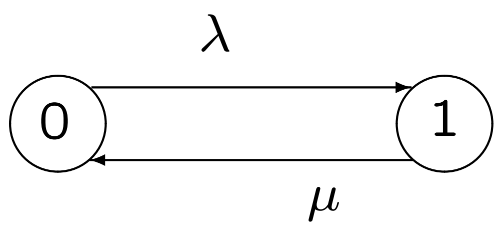
- For $n > 0$, this is

$$p_{n-1}\lambda_{n-1} + p_{n+1}\mu_{n+1} = p_n(\lambda_n + \mu_n)$$

- For $n = 0$, this is

$$p_1\mu_1 = p_0\lambda_0$$

Example 4.3.2 (Paradigm)



1. **Write down the balance equations.**

$$\begin{aligned} p_0\lambda &= p_1\mu \\ p_1\mu &= p_0\lambda \end{aligned}$$

2. **Write down the normalizing equation.**

$$p_0 + p_1 = 1.$$

3. **Solve for p_0 and p_1.**

$$\begin{aligned} p_0 &= \mu/(\lambda + \mu) \\ p_1 &= \lambda/(\lambda + \mu) \end{aligned}$$

4. **Interpret result.**

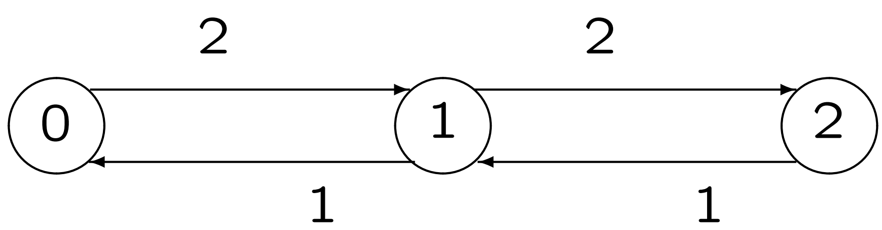
Notes for Example 4.3.2.

These four steps provide a general paradigm for finding the stationary probabilities of a birth-and-death process and interpreting its limiting behavior.

For this particular example, the interpretation is that the time spent in either state is proportional to the arrival rate into that state.

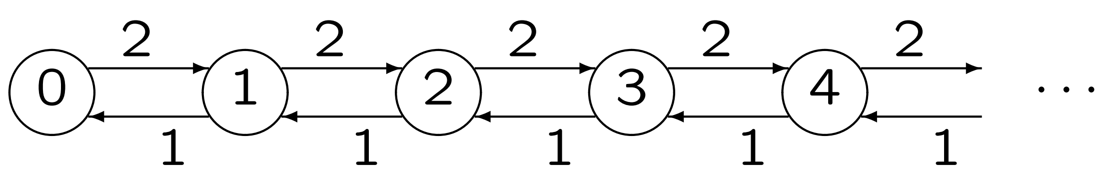
Exercise

Consider this 3-state birth-and-death process.



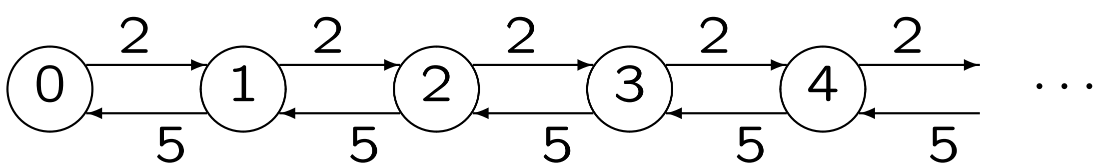
- Use the paradigm to find p_0, p_1, and p_2.

Consider this infinite-state birth-and-death process.



- In consideration of your answer for the process above, can you guess what the limiting probabilities for this process are?

Final Birth-and-Death Exercise



Balance Equations. For $n > 0$, we have

$$\begin{aligned} 5p_{n+1} + 2p_{n-1} &= 7p_n \\ 5p_1 &= 2p_0 \end{aligned}$$

Normalizing Equation.

$$\sum_{n=0}^{\infty} p_n = 1$$

Solving. Use your hard-earned knowledge of solving recurrences to solve for the p_n. The recurrence is

$$5p_{n+1} \quad = \quad 7p_n - 2p_{n-1}$$

for $n > 0$.

Interpretation. You can solve this kind of problem all day! ☺

Markov Chains Lite

Start with a discrete parameter, discrete state space, stochastic process

$$\mathsf{X} \quad = \quad \{X_n : n \in Z^+, X_n \in Z^+\}.$$

Suppose X is **homogeneous in time** (time is discrete). This means that, for all times $n_1, n_2 \in Z^+$, and all states $i, j \in Z$,

$$\begin{aligned} \Pr\left[X_{n_1+1} = j | X_{n_1} = i\right] &= \Pr\left[X_{n_2+1} = j | X_{n_2} = i\right] \\ &= P_{i,j} \end{aligned}$$

where $P_{i,j}$ is a constant independent of time.

Then we have a **Markov chain with stationary transition probabilities.**

Transition Probability Matrix

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots \\ P_{2,0} & P_{2,1} & P_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ P_{i,0} & P_{i,1} & P_{i,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

EXAMPLE.

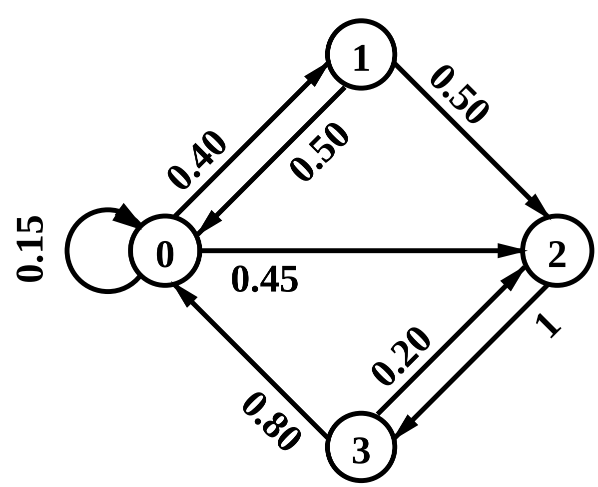
$$P = \begin{bmatrix} 0.15 & 0.40 & 0.45 & 0 \\ 0.50 & 0 & 0.50 & 0 \\ 0 & 0 & 0 & 1 \\ 0.80 & 0 & 0.20 & 0 \end{bmatrix}$$

Note that every row sums to 1.

Markov Chain Diagram

$$P = \begin{bmatrix} 0.15 & 0.40 & 0.45 & 0 \\ 0.50 & 0 & 0.50 & 0 \\ 0 & 0 & 0 & 1 \\ 0.80 & 0 & 0.20 & 0 \end{bmatrix}$$

Graph with states as nodes, arc from i to j labeled $P_{i,j}$ if nonzero.



Important Properties

- **Reachability.** If there is a path in the diagram from node i to node j, then j is **reachable** from i.

- **Irreducibility.** If every node is reachable from every other node, then the Markov chain is **irreducible.**

Multiple Steps

The n-step transition probability from node i to node j is

$$P_{i,j}^{(n)} \quad = \quad \Pr\left[X_n = j | X_0 = i\right].$$

- **2-step transitions.**

  $$P_{i,j}^{(2)} \quad = \quad \sum_{k=0}^{\infty} P_{i,k} P_{k,j}$$

  This is the inner product of the ith row and jth column of P. Said another way

  $$\left[\quad P_{i,j}^{(2)} \quad \right] \quad = \quad P^2.$$

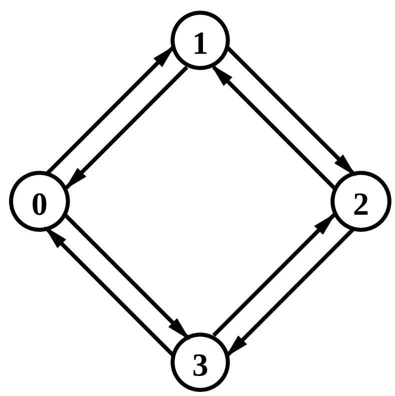
- ***n*-step transitions.** Generalize this observation to get

  $$\left[\quad P_{i,j}^{(n)} \quad \right] \quad = \quad P^n.$$

Period

The **period** of state i is the greatest common divisor of all integers $n \geq 1$ such that $P_{i,i}^{(n)} > 0$.

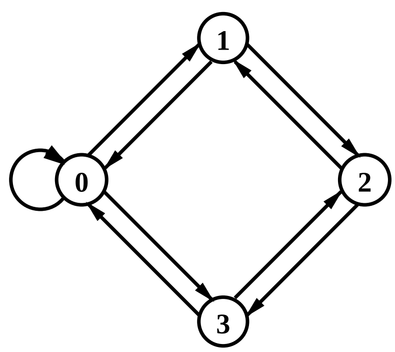
- **Periodic state.** Period is > 1.



Period of state 0 is ?

Period of state 1 is ?

- **Aperiodic state.** Period is 1.



Period of state 0 is ?

Period of state 1 is ?

Probability of Recurrence

The probability that the first recurrence of state i occurs after n steps is

$$f_i^{(n)} = \Pr\left[X_n = i, X_1 \neq i \wedge \cdots \wedge X_{n-1} \neq i \mid X_0 = i\right].$$

Accumulate these to get the probability of ever returning to state i:

$$f_i = \sum_{n=1}^{\infty} f_i^{(n)}.$$

- If $f_i < 1$, then state i is **transient.**
- If $f_i = 1$, then state i is **recurrent.**

Recurrent States

If state i is recurrent, that is,

$$\sum_{n=1}^{\infty} f_i^{(n)} \quad = \quad 1,$$

then define the **mean recurrence time of state** i to be

$$m_i \quad = \quad \sum_{n=1}^{\infty} n f_i^{(n)}.$$

- If $m_i = \infty$, then i is **recurrent null.**
- If $m_i < \infty$, then i is **positive recurrent.**

Stationary and Limiting Probabilities

The vector $\pi = (\pi_0, \pi_1, \ldots)$, where each $\pi_i \geq 0$ and $\sum_{n=0}^{\infty} \pi_n = 1$, is a **stationary probability distribution** for X if this matrix equation is satisfied:

$$\pi \quad = \quad \pi P.$$

Example of a Markov chain without a stationary probability distribution $\boxed{\quad ?}$

The vector $\pi = (\pi_0, \pi_1, \ldots)$ is a **limiting probability distribution** for X if for each i, $i \geq 0$,

$$\pi_i \quad = \quad \lim_{n \to \infty} \Pr [X_n = i] \,.$$

Example of a Markov chain without a limiting probability distribution $\boxed{\quad ?}$

Theorem 4.4.4

Suppose X is an irreducible, aperiodic, time homogeneous Markov chain.

- The limiting probabilities
$$\pi_i = \lim_{n\to\infty} \Pr[X_n = i]$$
always exist and are independent of the initial distribution.

- If all the states are **not** positive recurrent, then all $\pi_i = 0$ and no stationary probability distribution exists.

- **Ergodicity.** Otherwise, $\pi = (\pi_0, \pi_1, \ldots)$ is the stationary probability distribution, and hence π is the unique solution to
$$\begin{aligned} \sum_{i=0}^{\infty} \pi_i &= 1 \\ \pi &= \pi P. \end{aligned}$$

Paradigm

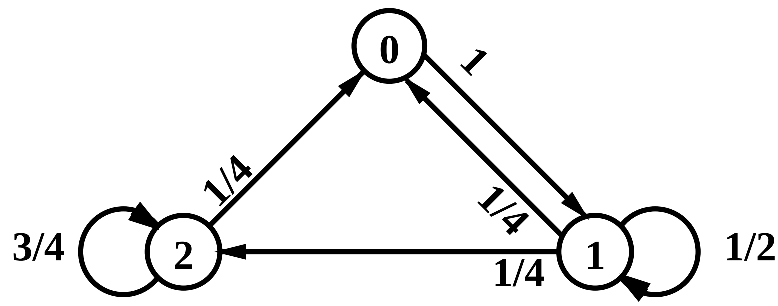
0. **Model your problem using a Markov chain.**

1. **Write down the equations in Theorem 4.4.4.**

2. **Solve the equations for the π_i's.**

3. **Interpret the results.**

On to an example!



1. **Write down the equations.**

$$
\begin{aligned}
\pi_0 + \pi_1 + \pi_2 &= 1 \\
\pi_0 &= \frac{1}{4}\pi_1 + \frac{1}{4}\pi_2 \\
\pi_1 &= 1 \cdot \pi_0 + \frac{1}{2}\pi_1 \\
\pi_2 &= \frac{1}{4}\pi_1 + \frac{3}{4}\pi_2
\end{aligned}
$$

2. **Solve the equations for the π_i's.**

$$
\begin{aligned}
\pi_0 &= \tfrac{1}{5} \\
\pi_1 &= \tfrac{2}{5} \\
\pi_2 &= \tfrac{2}{5}
\end{aligned}
$$

3. **Interpret the results.** $\boxed{\quad ?}$

Convergence of Markov Chains

- Given a Markov chain $\mathsf{X} = \{X_n : n \in Z^+, X_n \in Z^+\}$ with transition probability matrix P, it may have a stationary **and** limiting probability distribution

$$\pi = (\pi_0, \pi_1, \pi_2, \ldots).$$

  Then we know the following:

$$\sum_{n=0}^{\infty} \pi_n = 1 \quad (2)$$

$$\pi = \pi P \quad (3)$$

$$\pi_i = \lim_{n \to \infty} \Pr[X_n = i]. \quad (4)$$

- How rapid is the convergence?

Analyzing Convergence Rates

- The probability of being in state i at time n is

$$\pi_i^{(n)} \quad = \quad \Pr[X_n = i].$$

- $$\pi^{(n)} \quad = \quad \left(\pi_0^{(n)}, \pi_1^{(n)}, \ldots\right)$$

  is the vector of state probabilities at time n.

- The initial probability distribution is $\pi^{(0)}$.

Analyzing Convergence Rates (Continued)

- The **deviation** in state i from its limiting probability at time n is

$$\delta_i^{(n)} = \pi_i - \pi_i^{(n)}.$$

- The vector of deviations is

$$\delta^{(n)} = \left(\delta_0^{(n)}, \delta_1^{(n)}, \ldots\right).$$

- For all $n \in Z^+$, we get the vector equation

$$\pi = \pi^{(n)} + \delta^{(n)}.$$

- Also,

$$\sum_{i=0}^{\infty} \delta_i^{(n)} = 0.$$

Analyzing Convergence Rates (Continued)

- For all $n \in Z^+$, we get the recurrence

$$\pi^{(n+1)} = \pi^{(n)} P.$$

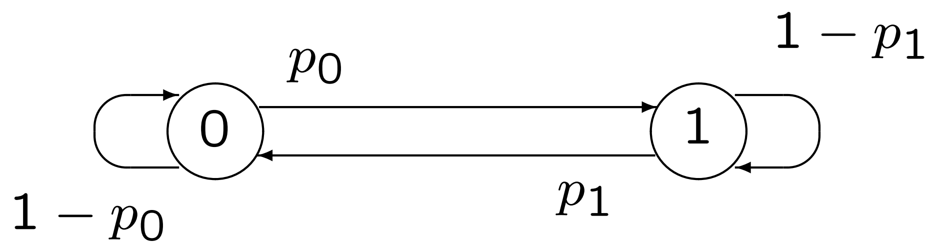
- By linearity of matrix multiplication, it follows that, for all n,

$$\delta^{(n+1)} = \delta^{(n)} P.$$

- Analyze the rate at which $\delta^{(n)} \rightarrow 0$ to determine how rapidly the Markov chain "mixes."

- Example next!

Convergence Example



- $$P = \begin{bmatrix} 1 - p_0 & p_0 \\ p_1 & 1 - p_1 \end{bmatrix}.$$

- As in the analysis of the similar birth-and-death process, the paradigm gives the limiting probabilities

$$\pi_0 = \frac{p_1}{p_0 + p_1}$$

$$\pi_1 = \frac{p_0}{p_0 + p_1}$$

Example Continued

- The recurrence gives

$$\begin{aligned} \delta_0^{(n+1)} &= (1-p_0)\delta_0^{(n)} + p_1\delta_1^{(n)} \\ \delta_1^{(n+1)} &= p_0\delta_0^{(n)} + (1-p_1)\delta_1^{(n)} \end{aligned}$$

- Use $\delta_0^{(n)} + \delta_1^{(n)} = 0$ to get

$$\begin{aligned} \delta_0^{(n+1)} &= (1-p_0)\delta_0^{(n)} - p_1\delta_0^{(n)} \\ &= (1-p_0-p_1)\,\delta_0^{(n)} \\ &= (1-p_0-p_1)^{n+1}\,\delta_0^{(0)} \end{aligned}$$

- Convergence is exponentially rapid.

- When fastest $\boxed{\quad ?}$ When slowest $\boxed{\quad ?}$