STAT 380: Introduction to Stochastic Processes

Course outline:

- Basic concepts of probability. Review of
  - Distributions
  - Expectation and moments
  - Moment generating functions
  - Independence, conditioning
- Markov Chains
- Poisson Processes
- Birth and Death Processes
- Brownian Motion
- Monte Carlo Simulation techniques
  - Random number generators
  - Generating random variables

Today's lecture Summary

- Three examples to review
  - basic probability: sample space, events, random variables, expected value
  - standard distributions: Binomial, Geometric, Poisson
  - conditional probability, independence
  - Bayes' rule
- Introduction to modelling

Some Basic Examples

Example 1: Three cards: one red on both sides, one black on both sides, one black on one side, red on the other. Shuffle, pick card at random. Side up is Black. What is the probability the side down is Black?

Solution: To do this carefully, enumerate *sample space*, Ω, of all possible outcomes. Six sides to the three cards. Label three red sides 1, 2, 3 with sides 1, 2 on the all red card (card # 1). Label three black sides 4, 5, 6 with 3, 4 on opposite sides of mixed card (card #2). Define some *events*:

$$\begin{aligned} A_i &= \{\text{side } i \text{ shows face up}\} \\ B &= \{\text{side showing is black}\} \\ C_j &= \{\text{card } j \text{ is chosen}\} \end{aligned}$$

One representation $\Omega = \{1,2,3,4,5,6\}$. Then $A_i = \{i\}$, $B = \{4,5,6\}$, $C_1 = \{1,2\}$ and so on.

Modelling: assumptions about some probabilities; deduce probabilities of other events. In example simplest model is

All of the A_i are equally likely.

Apply two rules:

$$P(\cup_1^6 A_i) = \sum_1^6 P(A_i) \quad \text{and} \quad P(\Omega) = 1$$

to get, for $i = 1, \ldots, 6$,

$$P(A_i) = \frac{1}{6}$$

Question was about down side of card. We have been told B has happened. Event that a black side is down is $D = \{3, 5, 6\}$. (Of course B has happened rules out 3.)

Definition of conditional probability:

$$P(D|B) = \frac{P(D \cap B)}{P(B)} = \frac{P(\{5,6\})}{P(\{4,5,6\})} = \frac{2}{3}$$

Example 2: Monte Hall, Let's Make a Deal. Monte shows you 3 doors. Prize hidden behind one. You pick a door. Monte opens a door you didn't pick; shows you no prize; offers to let you switch to the third door. Do you switch?

Sample space: typical element is (a, b, c) where a is number of door with prize, b is number of your first pick and c is door Monte opens with no prize.

(1,1,2)	(1,1,3)	(1,2,3)	(1,3,2)
(2,1,3)	(2,2,1)	(2,2,3)	(2,3,1)
(3,1,2)	(3,2,1)	(3,3,1)	(3,3,2)

Model? Traditionally we define events like

$$\begin{aligned} A_i &= \{\text{Prize behind door } i\} \\ B_j &= \{\text{You choose door } j\} \end{aligned}$$

and assume that each A_i has chance $1/3$. We are assuming we have no prior reason to suppose Monte favours one door. But these and all other probabilities depend on the behaviour of people so are open to discussion.

The event LS, that you lose if you switch is

$$(A_1 \cap B_1) \cup (A_2 \cap B_2) \cup (A_3 \cap B_3)$$

The natural modelling assumption, which captures the idea that you have no idea where the prize is hidden, is that each A_i is *independent* of each B_j, that is,

$$P(A_i \cap B_j) = P(A_i)P(B_j)$$

Usually we would phrase this assumption in terms of two *random variables*, M, the door with the prize, and C the door you choose. We are assuming that M and C are independent. Then

$$\begin{aligned} P(LS) =& P(A_1 \cap B_1) + P(A_2 \cap B_2) \\ &+ P(A_3 \cap B_3) \\ =& P(A_1)P(B_1) + P(A_2)P(B_2) \\ &+ P(A_3)P(B_3) \\ =& \frac{1}{3}\{P(B_1) + P(B_2) + P(B_3)\} \\ =& \frac{1}{3} \end{aligned}$$

So the event you win by switching has probability $2/3$ and you should switch.

Usual phrasing of problem. You pick 1, Monte shows 3. Should you take 2? Let S be rv $S =$ door Monte shows you. Question:

$$P(M = 1 | C = 1, S = 3)$$

Modelling assumptions *do not* determine this; it depends on Monte's method for choosing door to show when he has a choice. Two simple cases:

1. Monte picks at random so
$$P(S = 3|M = 1, C = 1) = 1/2$$
2. Monte chooses the door with the largest possible number:
$$P(S = 3|M = 1, C = 1) = 1$$

Use *Bayes' rule*:

$$P(M = 1|C = 1, S = 3) = \frac{P(M = 1, C = 1, S = 3)}{P(C = 1, S = 3)}$$

Numerator is

$$P(S = 3|M = 1, C = 1)P(M = 1, C = 1) = P(S = 3|M = 1, C = 1)P(C = 1)/3$$

Denominator is

$$P(S = 3|M = 1, C = 1)P(M = 1, C = 1) + P(S = 3|M = 2, C = 1)P(M = 2, C = 1) + P(S = 3|M = 3, C = 1)P(M = 3, C = 1)$$

which simplifies to

$$P(S = 3|M = 1, C = 1)P(M = 1)P(C = 1) + 1 \cdot P(M = 2)P(C = 1) + 0 \cdot P(M = 3)P(C = 1)$$

which in turn is
$$\{P(S = 3|M = 1, C = 1) + 1\} P(C = 1)/3$$

For case 1 we get
$$P(M = 1|C = 1, S = 3) = \frac{1/2}{1/2 + 1} = \frac{1}{3}$$
while for case 2 we get
$$P(M = 1|C = 1, S = 3) = \frac{1}{1 + 1} = \frac{1}{2}$$

Notice that in case 2 if we pick door 1 and Monte shows us door 2 we should *definitely* switch. Notice also that it would be normal to assume that Monte used the case 1 algorithm to pick the door to show when he has a choice; otherwise he is giving the contestant information. If the contestant knows Monte is using algorithm 2 then by switching if door 2 is shown and not if door 3 is shown he wins 2/3 of the time which is as good as the always switch strategy.

Example 3: Survival of family names. Traditionally: family name follows sons. Given man at end of 20th century. Probability descendant (male) with same last name alive at end of 21st century? or end of 30th century?

Simplified model: count generations not years. Compute probability, of survival of name for n generations.

Technically easier to compute q_n, probability of extinction by generation n.

Useful rvs:
$$X = \# \text{ of male children of first man}$$
$$Z_k = \# \text{ of male children in generation } k$$

Event of interest:

$$E_n = \{Z_n = 0\}$$

Break up E_n:

$$q_n = P(E_n) = \sum_{k=0}^{\infty} P(E_n \cap \{X = k\})$$

Now look at the event $E_n \cap \{X = k\}$. Let

$$\begin{aligned} B_{j,n-1} = &\{X = k\} \cap \{\text{child } j\text{s line extinct} \\ &\text{in } n-1 \text{ generations}\} \end{aligned}$$

Then

$$E_n \cap \{X = k\} = \cap_{j=1}^{k} B_{j,n-1}$$

Now add modelling assumptions:

1. Given (*conditional on*) $X = k$ the events $B_{j,n-1}$ are independent. In other words: one son's descendants don't affect other sons' descendants.
2. Given $X = k$ the probability of $B_{j,n-1}$ is q_{n-1}. In other words: sons are just like the parent.

Now add notation $P(X = k) = p_k$.

$$\begin{aligned} q_n &= \sum_{k=0}^{\infty} P(E_n \cap \{X = k\}) \\ &= \sum_{k=0}^{\infty} P(\cap_{j=1}^{k} B_{j,n-1} | X = k) p_k \\ &= \sum_{k=0}^{\infty} \prod_{1}^{k} P(B_{j,n-1} | X = k) p_k \\ &= \sum_{k=0}^{\infty} (q_{n-1})^k p_k \end{aligned}$$

Probability generating function:

$$\phi(s) = \sum_{k=0}^{\infty} s^k p_k = \mathrm{E}(s^X)$$

We have found

$$q_1 = p_0$$

and

$$q_n = \phi(q_{n-1})$$

Notice that $q_1 \le q_2 \le \cdots$ so that the limit of the q_n, say q_∞, must exist and (because ϕ is continuous) solve

$$q_\infty = \phi(q_\infty)$$

Special cases

Geometric Distribution: Assume

$$P(X = k) = (1-\theta)^k \theta \qquad k = 0, 1, 2, \ldots$$

(X is number of failures before first success. Trials are Bernoulli; θ is probability of success.)

Then

$$\begin{aligned}\phi(s) &= \sum_0^\infty s^k (1-\theta)^k \theta \\ &= \theta \sum_0^\infty [s(1-\theta)]^k \\ &= \frac{\theta}{1-s(1-\theta)}\end{aligned}$$

Set $\phi(s) = s$ to get

$$s[1-s(1-\theta)] = \theta$$

Two roots are

$$\frac{1 \pm \sqrt{1-4\theta(1-\theta)}}{2(1-\theta)} = \frac{1 \pm (1-2\theta)}{2(1-\theta)}$$

One of the roots is 1; the other is

$$\frac{\theta}{1-\theta}$$

If $\theta \ge 1/2$ the only root which is a probability is 1 and $q_\infty = 1$. If $\theta < 1/2$ then in fact $q_n \to q_\infty = \theta/(1-\theta)$.

Binomial(m, θ): If

$$P(X = k) = \binom{m}{k} \theta^k (1-\theta)^{m-k} \quad k = 0, \ldots, m$$

then

$$\begin{aligned}\phi(s) &= \sum_0^m \binom{m}{k} (s\theta)^k (1-\theta)^{m-k} \\ &= (1-\theta+s\theta)^m\end{aligned}$$

The equation $\phi(s) = s$ has two roots. One is 1. The other is less than 1 if and only if $m\theta = \mathrm{E}(X) > 1$.

Poisson(λ): Now

$$P(X = k) = e^{-\lambda}\lambda^k/k! \quad k = 0, 1, \ldots$$

and

$$\phi(s) = e^{\lambda(s-1)}$$

The equation $\phi(s) = s$ has two roots. One is 1. The other is less than 1 if and only if $\lambda = \mathrm{E}(X) > 1$.

Important Points:

- Use of conditioning.
- Approximate nature of modelling assumptions
- Key assumptions of conditional independence, *homogeneity*

Example 3: Mean values
Z_n = total number of sons in generation n.
$Z_0 = 1$ for convenience.
Compute $\mathrm{E}(Z_n)$.
Recall definition of expected value:
If X is discrete then

$$\mathrm{E}(X) = \sum_x x P(X = x)$$

If X is absolutely continuous then

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Theorem: If $Y = g(X)$, X has density f then

$$\mathrm{E}(Y) = \mathrm{E}(g(X)) = \int g(x) f(x) dx$$

Key properties of E:
1: If $X \geq 0$ then $\mathrm{E}(X) \geq 0$. Equals iff $P(X = 0) = 1$.
2: $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$.
3: If $0 \leq X_1 \leq X_2 \leq \cdots$ then

$$\mathrm{E}(\lim X_n) = \lim \mathrm{E}(X_n)$$

4: $\mathrm{E}(1) = 1$.

Conditional Expectations

If X, Y, two discrete random variables then

$$\mathrm{E}(Y|X = x) = \sum_y y P(Y = y|X = x)$$

Extension to absolutely continuous case:
Joint pmf of X and Y is defined as

$$p(x, y) = P(X = x, Y = y)$$

Notice: The pmf of X is

$$p_X(x) = \sum_y p(x, y)$$

Analogue for densities: joint density of X, Y is

$$f(x, y) dx dy \approx P(x \leq X \leq x + dx, y \leq Y \leq y + dy)$$

Interpretation is that

$$P(X \in A, Y \in B) = \int_A \int_B f(x, y) dy dx$$

Property: if X, Y have joint density $f(x, y)$ then X has density

$$f_X(x) = \int_y f(x, y) dy$$

Sums for discrete rvs are replaced by integrals.

Example:

$$f(x, y) = \begin{cases} x + y & 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

is a density because

$$\begin{aligned} \iint f(x, y) dx dy &= \int_0^1 \int_0^1 (x + y) dy dx \\ &= \int_0^1 x dx + \int_0^1 y dy \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

The marginal density of X is, for $0 \leq x \leq 1$.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ &= \int_0^1 (x + y) dy \\ &= (xy + y^2/2)\Big|_0^1 = x + \frac{1}{2} \end{aligned}$$

For x not in $[0,1]$ the integral is 0 so

$$f_X(x) = \begin{cases} x+\frac{1}{2} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Conditional Densities

If X and Y have joint density $f_{X,Y}(x,y)$ then we define the conditional density of Y given $X=x$ by analogy with our interpretation of densities. We take limits:

$$f_{Y|X}(y|x)dy \approx \frac{P(x \le X \le x+dx, y \le Y \le y+dy)}{P(x \le X \le x+dx)}$$

in the sense that if we divide through by dy and let dx and dy tend to 0 the conditional density is the limit

$$\frac{\lim_{dx,dy\to 0} \frac{P(x\le X\le x+dx, y\le Y\le y+dy)}{(dx\;dy)}}{\lim_{dx\to 0} \frac{P(x\le X\le x+dx)}{dx}}$$

Going back to our interpretation of joint densities and ordinary densities we see that our definition is just

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

When talking about a pair X and Y of random variables we refer to $f_{X,Y}$ as the *joint* density and to f_X as the *marginal* density of X.

Example: For f of previous example conditional density of Y given $X=x$ defined only for $0 \le x \le 1$:

$$f_{Y|X}(y|x) = \begin{cases} \frac{x+y}{x+\frac{1}{2}} & 0 \le x, y, \le 1 \\ 0 & 0 \le x \le 1, y > 1 \\ 0 & 0 \le x \le 1, y < 0 \\ \text{undefined} & otherwise \end{cases}$$

Example: X a Poisson(λ) random variable. Observe X then toss a coin X times. Y is number of heads. $P(H) = p$

$$\begin{aligned} f_Y(y) &= \sum_x f_{X,Y}(x,y) \\ &= \sum_x f_{Y|X}(y|x) f_X(x) \\ &= \sum_{x=0}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} \times \frac{\lambda^x}{x!} e^{-\lambda} \end{aligned}$$

WARNING: in sum $0 \le y \le x$ is required and x, y integers so sum really runs from y to ∞

$$\begin{aligned} f_Y(y) &= \frac{(p\lambda)^y e^{-\lambda}}{y!} \sum_{x=y}^{\infty} \frac{[(1-p)\lambda]^{x-y}}{(x-y)!} \\ &= \frac{(p\lambda)^y e^{-\lambda}}{y!} \sum_{0}^{\infty} \frac{[(1-p)\lambda]^k}{k!} \\ &= \frac{(p\lambda)^y e^{-\lambda}}{y!} e^{(1-p)\lambda} \\ &= e^{-p\lambda} (p\lambda)^y / y! \end{aligned}$$

which is a Poisson($p\lambda$) distribution.

Conditional Expectations

If X and Y are continuous random variables with joint density $f_{X,Y}$ we define:

$$E(Y|X=x) = \int y f_{Y|X}(y|x)dy$$

Key properties of conditional expectation

1: If $Y \geq 0$ then $\mathrm{E}(Y|X=x) \geq 0$. Equals iff $P(Y=0|X=x)=1$.

2: $\mathrm{E}(A(X)Y + B(X)Z|X=x) = A(x)\mathrm{E}(Y|X=x) + B(x)\mathrm{E}(Z|X=x)$.

3: If Y and X are independent then

$$\mathrm{E}(Y|X=x) = \mathrm{E}(Y)$$

4: $\mathrm{E}(1|X=x) = 1$.

Example:

$$f(x,y) = \begin{cases} x+y & 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

has conditional of $Y|X$:

$$f_{Y|X}(y|x) = \begin{cases} \frac{x+y}{x+\frac{1}{2}} & 0 \leq x, y, \leq 1 \\ 0 & 0 \leq x \leq 1, y > 1 \\ 0 & 0 \leq x \leq 1, y < 0 \\ \text{undefined} & otherwise \end{cases}$$

so, for $0 \leq x \leq 1$,

$$\begin{aligned} \mathrm{E}(Y|X=x) &= \int_0^1 y \frac{x+y}{x+\frac{1}{2}} dy \\ &= \frac{x/2 + 1/3}{x + 1/2} \end{aligned}$$

Computing expectations by conditioning:

Notation: $\mathrm{E}(Y|X)$ is the function of X you get by working out $\mathrm{E}(Y|X=x)$, getting a formula in x and replacing x by X. This makes $\mathrm{E}(Y|X)$ a random variable.

Properties:

1: $\mathrm{E}(A(X)Y + B(X)Z|X) = A(X)\mathrm{E}(Y|X) + B(X)\mathrm{E}(Z|X)$.

2: If Y and X are independent then

$$\mathrm{E}(Y|X) = \mathrm{E}(Y)$$

3: $\mathrm{E}(1|X) = 1$.

4: $\mathrm{E}\left[\mathrm{E}(Y|X)\right] = \mathrm{E}(Y)$ (compute average holding X fixed first, then average over X).

In example:

$$\mathrm{E}(Y|X) = \frac{X + 2/3}{2X+1}$$

Application to last names problem. Put $m = \mathrm{E}(X)$

$$
\begin{aligned}
\mathrm{E}(Z_n) &= \mathrm{E}\left[\mathrm{E}(Z_n|X)\right] \\
&= \mathrm{E}\left[X\mathrm{E}(Z_{n-1})\right] \\
&= \mathrm{E}(X)\mathrm{E}(Z_{n-1}) \\
&= m\mathrm{E}(Z_{n-1}) \\
&= m^2\mathrm{E}(Z_{n-2}) \\
&\ \ \vdots \\
&= m^{n-1}\mathrm{E}(Z_1) \\
&= m^n
\end{aligned}
$$

For $m < 1$ expect exponential decay. For $m > 1$ exponential growth (if not extinction).

Summary of Probability Review

We have reviewed the following definitions:

- **Probability Space** (or **Sample Space**): Ω = set of possible outcomes, ω.
- **Events**, subsets of Ω.
- **Probability**: P function defined for events, values in $[0,1]$ satisfying:
  1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
  2. **Countable additivity**: $A_1, A_2, \cdots$ **pairwise disjoint** ($j \neq k$ $A_j \cap A_k = \emptyset$)

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

- Joint, marginal, conditional pmfs: X, Y discrete:
  - Joint pmf
$$p_{X,Y}(x,y) = P(X = x, Y = y).$$
  - Marginal pmf
$$p_X(x) = P(X = x) = \sum_y p_{X,Y}(x,y)$$
  - Conditional pmf
$$p_{Y|X}(y|x) = p_{X,Y}(x,y)/p_X(x).$$
- Joint, marginal, conditional densities: X, Y continuous:
  - Joint density $f_{X,Y}(x,y)$; probabilities are double integrals
  - Marginal density
$$f_X(x) = \int f_{X,Y}(x,y)dy$$
  - Conditional density
$$f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x).$$
- Expected value:
$$\mathrm{E}(g(X)) = \sum g(x)p_X(x)$$
  or
$$\mathrm{E}(g(X)) = \int g(x)f_X(x)dx$$
- Conditional expectation
$$\mathrm{E}(g(Y)|X = x) = \sum g(y)p_{Y|X}(y|x)$$
  or
$$\mathrm{E}(g(Y)|X = x) = \int g(y)f_{Y|X}(y|x)dy$$

- $\mathrm{E}(g(Y)|X)$ is previous with x replaced by X after getting formula.

Tactics:

- Convert English expressions to set notation:
  - "Or" means union (remember *inclusive* or).
  - "And" means intersection.
  - "not" means complement.
- Compute prob something happens as 1-prob it doesn't.
- Break up event into disjoint pieces and add up probabilities.
- Find independent events and write event as intersection.
- Find A, B for which $P(A|B)$ is known.
- Use Bayes rule: if $B_1, \ldots, B_p$ is a *partition* then

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{\sum P(A|B_j)P(B_j)}$$

- Use first step analysis. (See family names example, craps game, alternating dice tossing on asst 1.)

Tactics for expected values:

- Use linearity:

$$\mathrm{E}(\sum a_j X_j) = \sum a_j \mathrm{E}(X_j)$$

- Condition on another variable and use

$$\mathrm{E}\left[\mathrm{E}(Y|X)\right] = \mathrm{E}(Y)$$

- Use first step analysis. (Special case of previous.)

Markov Chains

Last names example has following structure: if, at generation n there are m individuals then number of sons in next generation has distribution of sum of m independent copies of the random variable X which is number of sons I have. This distribution does not depend on n, only on value m of Z_n. Call Z_n a **Markov Chain**.

Ingredients of a Markov Chain:

- A **state space** S. S will be finite or countable in this course.
- A sequence $X_0, X_1, \ldots$ of random variables whose values are all in S.
- Matrix $\mathbf{P}$ with entries $P_{i,j}$ for $i, j \in S$. $\mathbf{P}$ is required to be **stochastic**; for all i:

$$\sum_j P_{ij} = 1 \qquad \text{and} \qquad 0 \leq P_{ij}.$$

The **stochastic process** $X_0, X_1, \ldots$ is called a Markov chain if

$$P\left(X_{k+1} = j | X_k = i, A\right) = P_{i,j}$$

for *any* A defined in terms of $X_0, \ldots, X_{k-1}$ and for all i, j, k. Usually used with

$$A = \{X_{k-1} = i_{k-1}, \ldots, X_0 = i_0\}$$

for some $i_0, \ldots, i_{k-1}$.

The matrix $\mathbf{P}$ is called a *transition* matrix.

Example: If X in the last names example has a Poisson(λ) distribution then given $Z_n = k$, Z_{n+1} is like sum of k independent Poisson(λ) rvs which has a Poisson($k\lambda$) distribution. So

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ e^{-\lambda} & \lambda e^{-\lambda} & \lambda^2 e^{-\lambda}/2 & \cdots \\ e^{-2\lambda} & (2\lambda)e^{-2\lambda} & \frac{1}{2}(2\lambda)^2 e^{-2\lambda} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Example: Weather: each day is dry (D) or wet (W).

X_n is weather on day n.

Suppose dry days tend to be followed by dry days, say 3 times in 5 and wet days by wet 4 times in 5.

Markov assumption: yesterday's weather irrelevant to prediction of tomorrow's given today's.

Transition Matrix:

$$\mathbf{P} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

Now suppose it's wet today. P(wet in 2 days)?

$$P(X_2 = W | X_0 = W)$$

$$\begin{aligned}
=& P(X_2 = W, X_1 = D | X_0 = W) \\
&+ P(X_2 = W, X_1 = W | X_0 = W) \\
=& P(X_2 = W | X_1 = D, X_0 = W) \\
&\times P(X_1 = D | X_0 = W) \\
&+ P(X_2 = W | X_1 = W, X_0 = W) \\
&\times P(X_1 = W | X_0 = W) \\
=& P(X_2 = W | X_1 = D) \\
&\times P(X_1 = D | X_0 = W) \\
&+ P(X_2 = W | X_1 = W) \\
&\times P(X_1 = W | X_0 = W) \\
=& P_{W,D} P_{D,W} + P_{W,W} P_{W,W} \\
=& \left(\frac{1}{5}\right)\left(\frac{2}{5}\right) + \left(\frac{4}{5}\right)\left(\frac{4}{5}\right)
\end{aligned}$$

Notice that all the entries in the last line are items in $\mathbf{P}$. Look at the matrix product $\mathbf{PP}$:

$$\begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{11}{25} & \frac{14}{25} \\ \\ \frac{7}{25} & \frac{18}{25} \end{bmatrix} =$$

Notice that $P(X_2 = W | X_0 = W)$ is exactly the W, W entry in $\mathbf{PP}$.

General case. Define

$$P_{i,j}^{(n)} = P(X_n = j | X_0 = i)$$

Then

$$\begin{aligned} P(X_{m+n} = j | X_m = i, X_{m-1} = i_{m-1}, \ldots) \\ &= P(X_{m+n} = j | X_m = i) \\ &= P(X_n = j | X_0 = i) \\ &= P_{i,j}^{(n)} \\ &= (\mathbf{P}^n)_{i,j} . \end{aligned}$$

Proof of these assertions by induction on m, n.

Example for $n = 2$. Two bits to do:

First suppose U, V, X, Y are discrete variables. Assume

$$\begin{aligned} P(Y = y | X = x, U = u, V = v) \\ &= P(Y = y | X = x) \end{aligned}$$

for any x, y, u, v. Then I claim

$$\begin{aligned} P(Y = y | X = x, U = u) \\ &= P(Y = y | X = x) \end{aligned}$$

In words, if knowing both U and V doesn't change the conditional probability then knowing U alone doesn't change the conditional probability.

Proof of claim:

$$A = \{X = x, U = u\}$$

Then

$$\begin{aligned} P(Y = y | X = x, U = u) \\ &= \frac{P(Y = y, A)}{P(A)} \\ &= \frac{\sum_v P(Y = y, A, V = v)}{P(A)} \\ &= \frac{\sum_v P(Y = y | A, V = v) P(A, V = v)}{P(A)} \\ &= \frac{\sum_v P(Y = y | X = x) P(A, V = v)}{P(A)} \\ &= \frac{P(Y = y | X = x) \sum_v P(A, V = v)}{P(A)} \\ &= \frac{P(Y = y | X = x) P(A)}{P(A)} \\ &= P(Y = y | X = x) \end{aligned}$$

Second step: consider

$$\begin{aligned}
P(X_{n+2} = k|X_n = i) \\
&= \sum_j P(X_{n+2} = k, X_{n+1} = j|X_n = i) \\
&= \sum_j P(X_{n+2} = k|X_{n+1} = j, X_n = i) \\
&\qquad \times P(X_{n+1} = j|X_n = i) \\
&= \sum_j P(X_{n+2} = k|X_{n+1} = j) \\
&\qquad \times P(X_{n+1} = j|X_n = i) \\
&= \sum_j \mathbf{P}_{i,j}\mathbf{P}_{j,k}
\end{aligned}$$

This shows that

$$P(X_{n+2} = k|X_n = i) = (\mathbf{P}^2)_{i,k}$$

where $\mathbf{P}^2$ means the matrix product $\mathbf{PP}$. Notice both that the quantity does not depend on n and that we can compute it by taking a power of $\mathbf{P}$.

More general version

$$P(X_{n+m} = k|X_n = j) = (\mathbf{P}^m)_{j,k}$$

Since $\mathbf{P}^n\mathbf{P}^m = \mathbf{P}^{n+m}$ we get the Chapman-Kolmogorov equations:

$$P(X_{n+m} = k|X_0 = i) = \sum_j P(X_{n+m} = k|X_n = j)P(X_n = j|X_0 = i)$$

Summary: A Markov Chain has *stationary* n step transition probabilities which are the nth power of the 1 step transition probabilities.

Here is Maple output for the 1,2,4,8 and 16 step transition matrices for our rainfall example:

```
> p:= matrix(2,2,[[3/5,2/5],[1/5,4/5]]);
                         [3/5    2/5]
                    p := [          ]
                         [1/5    4/5]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> p16:=evalm(p8*p8):
```

This computes the powers (`evalm` understands matrix algebra).

Fact:

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

```
> evalf(evalm(p));
             [.6000000000    .4000000000]
             [                          ]
             [.2000000000    .8000000000]
> evalf(evalm(p2));
             [.4400000000    .5600000000]
             [                          ]
```

```
           [.2800000000    .7200000000]
> evalf(evalm(p4));
           [.3504000000    .6496000000]
           [                          ]
           [.3248000000    .6752000000]
> evalf(evalm(p8));
           [.3337702400    .6662297600]
           [                          ]
           [.3331148800    .6668851200]
> evalf(evalm(p16));
           [.3333336197    .6666663803]
           [                          ]
           [.3333331902    .6666668098]
```

Where did 1/3 and 2/3 come from?

Suppose we toss a coin $P(H) = \alpha_D$ and start the chain with Dry if we get heads and Wet if we get tails. Then

$$P(X_0 = x) = \begin{cases} \alpha_D & x = \text{Dry} \\ \alpha_W = 1 - \alpha_D & x = \text{Wet} \end{cases}$$

and

$$\begin{aligned} P(X_1 = x) &= \sum_y P(X_1 = x | X_0 = y) P(X_0 = y) \\ &= \sum_y \alpha_y P_{y,x} \end{aligned}$$

Notice last line is a matrix multiplication of row vector α by matrix $\mathbf{P}$. A special α: if we put $\alpha_D = 1/3$ and $\alpha_W = 2/3$ then

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

In other words if we start off $P(X_0 = D) = 1/3$ then $P(X_1 = D) = 1/3$ and analogously for W. This means that X_0 and X_1 have the same distribution.

A probability vector α is called an initial distribution for the chain if

$$P(X_0 = i) = \alpha_i$$

A Markov Chain is **stationary** if

$$P(X_1 = i) = P(X_0 = i)$$

for all i

An initial distribution is called **stationary** if the chain is stationary. We find that α is a stationary initial distribution if

$$\alpha \mathbf{P} = \alpha$$

Suppose $\mathbf{P}^n$ converges to some matrix $\mathbf{P}^\infty$. Notice that

$$\lim_{n\to\infty} \mathbf{P}^{n-1} = \mathbf{P}^\infty$$

and

$$\begin{aligned} \mathbf{P}^\infty &= \lim \mathbf{P}^n \\ &= \left[\lim \mathbf{P}^{n-1}\right] \mathbf{P} \\ &= \mathbf{P}^\infty \mathbf{P} \end{aligned}$$

This proves that each row α of $\mathbf{P}^\infty$ satisfies

$$\alpha = \alpha \mathbf{P}$$

Def'n: A row vector x is a left eigenvector of A with eigenvalue λ if

$$xA = \lambda x$$

So each row of $\mathbf{P}^\infty$ is a left eigenvector of $\mathbf{P}$ with eigenvalue 1.

Finding stationary initial distributions. Consider the $\mathbf{P}$ for the weather example. The equation

$$\alpha \mathbf{P} = \alpha$$

is really

$$\begin{aligned} \alpha_D &= 3\alpha_D/5 + \alpha_W/5 \\ \alpha_W &= 2\alpha_D/5 + 4\alpha_W/5 \end{aligned}$$

The first can be rearranged to

$$\alpha_W = 2\alpha_D$$

and so can the second. If α is to be a probability vector then

$$\alpha_W + \alpha_D = 1$$

so we get

$$1 - \alpha_D = 2\alpha_D$$

leading to

$$\alpha_D = 1/3$$

Some more examples:

$$\mathbf{P} = \begin{bmatrix} 0 & 1/3 & 0 & 2/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 2/3 & 0 & 1/3 & 0 \end{bmatrix}$$

Set $\alpha \mathbf{P} = \alpha$ and get

$$\begin{aligned} \alpha_1 &= \alpha_2/3 + 2\alpha_4/3 \\ \alpha_2 &= \alpha_1/3 + 2\alpha_3/3 \\ \alpha_3 &= 2\alpha_2/3 + \alpha_4/3 \\ \alpha_4 &= 2\alpha_1/3 + \alpha_3/3 \\ 1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

First plus third gives

$$\alpha_1 + \alpha_3 = \alpha_2 + \alpha_4$$

so both sums 1/2. Continue algebra to get

$$(1/4, 1/4, 1/4, 1/4).$$

```
p:=matrix([[0,1/3,0,2/3],[1/3,0,2/3,0],
          [0,2/3,0,1/3],[2/3,0,1/3,0]]);

               [ 0      1/3      0      2/3]
               [                           ]
               [1/3      0      2/3      0 ]
```

```
          p := [                         ]
               [ 0     2/3     0     1/3]
               [                         ]
               [2/3     0     1/3     0 ]
> p2:=evalm(p*p);
             [5/9     0     4/9     0 ]
             [                        ]
             [ 0     5/9     0     4/9]
        p2:= [                        ]
             [4/9     0     5/9     0 ]
             [                        ]
             [ 0     4/9     0     5/9]
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> p16:=evalm(p8*p8):
> p17:=evalm(p8*p8*p):
> evalf(evalm(p16));
    [.5000000116 , 0 , .4999999884 , 0]
    [                                 ]
    [0 , .5000000116 , 0 , .4999999884]
    [                                 ]
    [.4999999884 , 0 , .5000000116 , 0]
    [                                 ]
    [0 , .4999999884 , 0 , .5000000116]
> evalf(evalm(p17));
    [0 , .4999999961 , 0 , .5000000039]
    [                                 ]
    [.4999999961 , 0 , .5000000039 , 0]
    [                                 ]
    [0 , .5000000039 , 0 , .4999999961]
    [                                 ]
    [.5000000039 , 0 , .4999999961 , 0]
> evalf(evalm((p16+p17)/2));
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
```

$\mathbf{P}^n$ doesn't converges but$(\mathbf{P}^n + \mathbf{P}^{n+1})/2$ does. Next example:

$$p = \begin{bmatrix} \frac{2}{5} & \frac{3}{5} & 0 & 0 \\ \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

Solve $\alpha \mathbf{P} = \alpha$:

$$\begin{aligned} \alpha_1 &= \frac{2}{5}\alpha_1 + \frac{1}{5}\alpha_2 \\ \alpha_2 &= \frac{3}{5}\alpha_1 + \frac{4}{5}\alpha_2 \\ \alpha_3 &= \frac{2}{5}\alpha_3 + \frac{1}{5}\alpha_4 \\ \alpha_4 &= \frac{3}{5}\alpha_3 + \frac{4}{5}\alpha_4 \\ 1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

Second and fourth equations redundant. Get

$$\begin{aligned}\alpha_2 &= 3\alpha_1\\ 3\alpha_3 &= \alpha_4\\ 1 &= 4\alpha_1 + 4\alpha_3\end{aligned}$$

Pick α_1 in $[0, 1/4]$; put $\alpha_3 = 1/4 - \alpha_1$.

$$\alpha = (\alpha_1, 3\alpha_1, 1/4 - \alpha_1, 3(1/4 - \alpha_1))$$

solves $\alpha\mathbf{P} = \alpha$. So solution is not unique.

```
> p:=matrix([[2/5,3/5,0,0],[1/5,4/5,0,0],
           [0,0,2/5,3/5],[0,0,1/5,4/5]]);

                 [2/5    3/5     0      0 ]
                 [                        ]
                 [1/5    4/5     0      0 ]
            p := [                        ]
                 [ 0      0     2/5    3/5]
                 [                        ]
                 [ 0      0     1/5    4/5]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):

> evalf(evalm(p8*p8));
        [.2500000000 , .7500000000 , 0 , 0]
        [                                  ]
        [.2500000000 , .7500000000 , 0 , 0]
        [                                  ]
        [0 , 0 , .2500000000 , .7500000000]
        [                                  ]
        [0 , 0 , .2500000000 , .7500000000]
```

Notice that rows converge but to two different vectors:

$$\alpha^{(1)} = (1/4, 3/4, 0, 0)$$

and

$$\alpha^{(2)} = (0, 0, 1/4, 3/4)$$

Solutions of $\alpha\mathbf{P} = \alpha$ revisited? Check that

$$\alpha^{(1)}\mathbf{P} = \alpha^{(1)}$$

and

$$\alpha^{(2)}\mathbf{P} = \alpha^{(2)}$$

If $\alpha = \lambda\alpha^{(1)} + (1-\lambda)\alpha^{(2)}$ $(0 \le \lambda \le 1)$ then

$$\alpha\mathbf{P} = \alpha$$

so again solution is not unique. Last example:

```
> p:=matrix([[2/5,3/5,0],[1/5,4/5,0],
            [1/2,0,1/2]]);

                    [2/5    3/5     0 ]
```

```
                [                  ]
           p := [1/5    4/5      0 ]
                [                  ]
                [1/2     0     1/2]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> evalf(evalm(p8*p8));
  [.2500000000 .7500000000        0       ]
  [                                       ]
  [.2500000000 .7500000000        0       ]
  [                                       ]
  [.2500152588 .7499694824 .00001525878906]
```

Interpretation of examples

- For some $\mathbf{P}$ all rows converge to some α. In this case this α is a stationary initial distribution.
- For some $\mathbf{P}$ the locations of zeros flip flop. $\mathbf{P}^n$ does not converge. Observation: average

$$\frac{\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^n}{n}$$

  does converge.
- For some $\mathbf{P}$ some rows converge to one α and some to another. In this case the solution of $\alpha\mathbf{P}=\alpha$ is not unique.

Basic distinguishing features: pattern of 0s in matrix $\mathbf{P}$.

Classification of States

State i **leads to** state j if $\mathbf{P}_{ij}^n > 0$ for some n. It is convenient to agree that $\mathbf{P}^0 = \mathbf{I}$ the identity matrix; thus i leads to i.

Note i leads to j and j leads to k implies i leads to k (Chapman-Kolmogorov).

States i and j **communicate** if i leads to j and j leads to i.

The relation of communication is an **equivalence** relation; it is reflexive, symmetric and transitive: if i and j communicate and j and k communicate then i and k communicate.

Example (+ signs indicate non-zero entries):

$$\mathbf{P}=\left[\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ + & + & + & 0 & 0 \\ + & 0 & 0 & + & 0 \\ 0 & + & 0 & + & + \end{array}\right]$$

For this example:
$1 \rightsquigarrow 2$, $2 \rightsquigarrow 3$, $3 \rightsquigarrow 1$ so $1,2,3$ are all in the same communicating class.
$4 \rightsquigarrow 1,2,3$ but not vice versa.
$5 \rightsquigarrow 1,2,3,4$ but not vice versa.
So the communicating classes are

$$\{1,2,3\} \quad \{4\} \quad \{5\}$$

A Markov Chain is **irreducible** if there is only one communicating class.
Notation:

$$f_i = P(\exists n > 0 : X_n = i | X_0 = i)$$

State i is **recurrent** if $f_i = 1$, otherwise **transient**.

If $f_i = 1$ then Markov property (chain starts over when it gets back to i) means prob return infinitely many times (given started in i or given ever get to i) is 1.

Consider chain started from transient i. Let N be number of visits to state i (including visit at time 0). To return m times must return once then starting over return $m-1$ times, then never return. So:

$$P(N = m|X_0 = i) = f_i^{m-1}(1 - f_i)$$

for $m = 1, 2, \ldots$.

N has a Geometric distribution and $\mathrm{E}(N|X_0 = i) = 1/(1 - f_i)$.

Another calculation:

$$N = \sum_{k=0}^{\infty} 1(X_k = i)$$

so

$$\mathrm{E}(N|X_0 = i) = \sum_{k=0} 0^{\infty} P(X_k = i)$$

If we start the chain in state i then this is

$$\mathrm{E}(N|X_0 = i) = \sum_{k=0} 0^{\infty} \mathbf{P}_{ii}^k$$

and i is transient if and only if

$$\sum_{k=0} 0^{\infty} \mathbf{P}_{ii}^k < \infty\,.$$

For last example: 4 and 5 are transient. Claim: states 1, 2 and 3 are recurrent.

Proof: argument above shows each transient state is visited only finitely many times. So: there is a recurrent state. (Note use of finite number of states.) It must be one of 1, 2 and 3. **Proposition**: If one state in a communicating class is recurrent then all states in the communicating class are recurrent.

Proof: Let i be the known recurrent state so

$$\sum_n \mathbf{P}_{ii}^n = \infty$$

Assume i and j communicate. Find integers m and k such that

$$\mathbf{P}_{ij}^m > 0$$

and

$$\mathbf{P}_{ji}^k > 0$$

Then

$$\mathbf{P}_{jj}^{m+n+k} \ge \mathbf{P}_{ji}^k \mathbf{P}_{ii}^n \mathbf{P}_{ij}^m$$

Sum RHS over n get ∞ so

$$\sum_n \mathbf{P}_{jj}^n = \infty$$

Proposition also means that if 1 state in a class is transient so are all.

State i has period d if d is greatest common divisor of

$$\{n : \mathbf{P}_{ii}^n > 0\}$$

If i and j are in the same class then i and j have same period. If $d = 1$ then state i is called **aperiodic**; if $d > 1$ then i is periodic.

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

For this example $\{1, 2, 3\}$ is a class of period 3 states and $\{4, 5\}$ a class of period 2 states.

$$\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

has a single communicating class of period 2.

A chain is **aperiodic** if all its states are aperiodic.

Infinite State Spaces

Example: consider a sequence of independent coin tosses with probability p of Heads on a single toss. Let X_n be number of heads minus number of tails after n tosses. Put $X_0 = 0$.

X_n is a Markov Chain. State space is $\mathbb{Z}$, the integers and

$$\mathbf{P}_{ij} = \begin{cases} p & j = i + 1 \\ 1 - p & j = i - 1 \\ 0 & \text{otherwise} \end{cases}$$

This chain has one communicating class (for $p \neq 0, 1$) and all states have period 2. According to the strong law of large numbers X_n/n converges to $2p - 1$. If $p \neq 1/2$ this guarantees that for all large enough n $X_n \neq 0$, that is, the number of returns to 0 is not infinite. The state 0 is then transient and so **all** states must be transient.

For $p = 1/2$ the situation is different. It is a fact that

$$\mathbf{P}_{00}^n = P(\#\text{ H} = \#\text{ T at time } n)$$

For n even this is the probability of exactly $n/2$ heads in n tosses.

Local Central Limit Theorem (normal approximation to $P(-1/2 < X_n < 1/2)$) (or Stirling's approximation) shows

$$\sqrt{2m}P(\text{Binomial}(2m, 1/2) = m) \to (2/\pi)^{1/2}$$

so:

$$\sum_n \mathbf{P}_{00}^n = \infty$$

That is: 0 is a recurrent state.

Hitting Times

Start irreducible recurrent chain X_n in state i. Let T_j be first $n > 0$ such that $X_n = j$. Define

$$m_{ij} = \mathrm{E}(T_j | X_0 = i)$$

First step analysis:

$$\begin{aligned} m_{ij} &= 1 \cdot P(X_1 = j | X_0 = i) \\ &\quad + \sum_{k \neq j} (1 + \mathrm{E}(T_j | X_0 = k)) P_{ik} \\ &= \sum_k P_{ik} + \sum_{k \neq j} P_{ik} m_{kj} \\ &= 1 + \sum_{k \neq j} P_{ik} m_{kj} \end{aligned}$$

Example

$$\mathbf{P} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

The equations are

$$m_{11} = 1 + \frac{2}{5} m_{21}$$
$$m_{12} = 1 + \frac{3}{5} m_{12}$$
$$m_{21} = 1 + \frac{4}{5} m_{21}$$
$$m_{22} = 1 + \frac{1}{5} m_{12}$$

The second and third equations give immediately

$$m_{12} = \frac{5}{2}$$
$$m_{21} = 5$$

Then plug in to the others to get

$$m_{11} = 3$$
$$m_{22} = \frac{3}{2}$$

Notice stationary initial distribution is

$$\left(\frac{1}{m_{11}}, \frac{1}{m_{22}}\right)$$

Consider fraction of time spent in state j:

$$\frac{1(X_0 = j) + \cdots + 1(X_n = j)}{n+1}$$

Imagine chain starts in state i; take expected value.

$$\frac{\sum_{r=1}^n \mathbf{P}_{ij}^r + 1(i=j)}{n+1} = \frac{\sum_{r=0}^n \mathbf{P}_{ij}^r}{n+1}$$

If rows of $\mathbf{P}^n$ converge to π then fraction converges to π_j; i.e. limiting fraction of time in state j is π_j.

Heuristic: start chain in i. Expect to return to i every m_{ii} time units. So are in state i about once every m_{ii} time units; i.e. limiting fraction of time in state i is $1/m_{ii}$.

Conclusion: for an irreducible recurrent finite state space Markov chain

$$\pi_i = \frac{1}{m_{ii}}.$$

Infinite State Spaces

Previous conclusion is still right **if** there is a stationary initial distribution.

Example: X_n = Heads − Tails after n tosses of fair coin. Equations are

$$m_{0,0} = 1 + \frac{1}{2} m_{1,0} + \frac{1}{2} m_{-1,0}$$
$$m_{1,0} = 1 + \frac{1}{2} m_{2,0}$$

and many more.

Some observations:

You have to go through 1 to get to 0 from 2 so

$$m_{2,0} = m_{2,1} + m_{1,0}$$

Symmetry (switching H and T):

$$m_{1,0} = m_{-1,0}$$

The transition probabilities are **homogeneous**:

$$m_{2,1} = m_{1,0}$$

Conclusion:

$$\begin{aligned} m_{0,0} &= 1 + m_{1,0} \\ &= 1 + 1 + \frac{1}{2} m_{2,0} \\ &= 2 + m_{1,0} \end{aligned}$$

Notice that there are **no** finite solutions!

Summary of the situation:

Every state is recurrent.

All the expected hitting times m_{ij} are infinite.

All entries $\mathbf{P}^n_{ij}$ converge to 0.

Jargon: The states in this chain are null recurrent.

One Example

Page 229, question 21: runner goes from front or back door, prob 1/2 each. Returns front or back, prob 1/2 each. Has k pairs of shoes, wears pair if any at departure door, leaves at return door. No shoes? Barefoot. Long run fraction of time barefoot?

Solution: Let X_n be number of shoes at front door on day n. Then X_n is a Markov Chain.

Transition probabilities:

k pairs at front door on day n. X_{n+1} is k if goes out back door (prob is 1/2) or out front door and back in front door (prob is 1/4). Otherwise X_{n+1} is $k-1$.

$0 < j < k$ pairs at front door on day n. X_{n+1} is $j+1$ if out back, in front (prob is 1/4). X_{n+1} is $j-1$ if out front, in back. Otherwise X_{n+1} is j.

0 pairs at front door on day n. X_{n+1} is 0 if out front door (prob 1/2) or out back door and in back door (prob 1/4) otherwise X_{n+1} is 1.

Transition matrix $\mathbf{P}$:

$$\begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & \cdots & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & \cdots & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

Doubly stochastic: row sums and column sums are 1.

So $\pi_i = 1/(k+1)$ for all i is stationary initial distribution.

Solution to problem: 1 day in $k+1$ no shoes at front door. Half of those go barefoot. Also 1 day in $k+1$ all shoes at front door; go barefoot half of these days. Overall go barefoot $1/(k+1)$ of the time.

Gambler's Ruin

Insurance company's reserves fluctuate: sometimes up, sometimes down. Ruin is event they hit 0 (company goes bankrupt). General problem. For given model of fluctuation compute probability of ruin either eventually or in next k time units.

Simplest model: gambling on Red at Casino. Bet \$1 at a time. Win \$1 with probability p, lose \$1 with probability $1-p$. Start with k dollars. Quit playing when down to \$0 or up to N. Compute

$$P_k = P(\text{reach } N \text{ before } 0|X_0 = k)$$

X_n = fortune after n plays. $X_0 = k$.
Transition matrix:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1-p & 0 & p & 0 & \cdots & 0 \\ 0 & 1-p & 0 & p & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

First step analysis:

$$\begin{aligned} P_0 &= 0 \\ P_i &= (1-p)P_{i-1} + pP_{i+1} \\ P_N &= 1 \end{aligned}$$

Middle equation is

$$pP_i + (1-p)P_i = (1-p)P_{i-1} + pP_{i+1}$$

or

$$\begin{aligned} P_{i+1} - P_i &= \frac{1-p}{p}(P_i - P_{i-1}) \\ &= \left(\frac{1-p}{p}\right)^2 (P_{i-1} - P_{i-2}) \\ &\vdots \\ &= \left(\frac{1-p}{p}\right)^i (P_1 - P_0) \\ &= \left(\frac{1-p}{p}\right)^i P_1 \end{aligned}$$

Sum from $i=0$ to $i=k-1$ to get

$$P_k = \sum_{i=0}^{k-1} \left(\frac{1-p}{p}\right)^i P_1$$

or

$$P_k = \frac{1 - \{(1-p)/p\}^k}{1 - \{(1-p)/p\}} P_1$$

For $k = N$ we get

$$1 = \frac{1 - \{(1-p)/p\}^N}{1 - \{(1-p)/p\}} P_1$$

so that

$$P_k = \frac{1 - \{(1-p)/p\}^k}{1 - \{(1-p)/p\}^N}$$

Notice that if $p = 1/2$ our formulas for the sum of the geometric series are wrong. But for $p = 1/2$ we get

$$P_k = kP_1$$

so

$$P_k = \frac{k}{N}.$$

Mean time in transient states

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{3}{8} & \frac{1}{8} \end{bmatrix}$$

States 3 and 4 are transient. Let $m_{i,j}$ be the expected total number of visits to state j for chain started in i.

For $i = 1$ or $i = 2$ and $j = 3$ or 4:

$$m_{ij} = 0$$

For $j = 1$ or $j = 2$

$$m_{ij} = \infty$$

For $i, j \in \{3, 4\}$ first step analysis:

$$\begin{aligned} m_{3,3} &= 1 + \frac{1}{4} m_{3,3} + \frac{1}{4} m_{4,3} \\ m_{3,4} &= 0 + \frac{1}{4} m_{3,4} + \frac{1}{4} m_{4,4} \\ m_{4,3} &= 0 + \frac{3}{8} m_{3,3} + \frac{1}{8} m_{4,3} \\ m_{4,4} &= 1 + \frac{3}{8} m_{3,4} + \frac{1}{8} m_{4,4} \end{aligned}$$

In matrix form

$$\begin{bmatrix} m_{3,3} & m_{3,4} \\ m_{4,3} & m_{4,4} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} \end{bmatrix} \begin{bmatrix} m_{3,3} & m_{3,4} \\ m_{4,3} & m_{4,4} \end{bmatrix}$$

Translate to matrix notation:

$$\mathbf{M} = \mathbf{I} + \mathbf{P}_T \mathbf{M}$$

where $\mathbf{I}$ is the identity, $\mathbf{M}$ is the matrix of means and $\mathbf{P}_T$ the part of the transition matrix corresponding to transient states.

Solution is

$$\mathbf{M} = (\mathbf{I} - \mathbf{P}_T)^{-1}$$

In our case

$$\mathbf{I} - \mathbf{P}_T = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} \\ -\frac{3}{8} & \frac{7}{8} \end{bmatrix}$$

so that

$$\mathbf{M} = \begin{bmatrix} \frac{14}{9} & \frac{4}{9} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

Poisson Processes

Particles arriving over time at a particle detector. Several ways to describe most common model.

Approach 1: numbers of particles arriving in an interval has Poisson distribution, mean proportional to length of interval, numbers in several non-overlapping intervals independent.

For $s < t$, denote number of arrivals in $(s, t]$ by $N(s, t)$. Model is

1. $N(s, t)$ has a Poisson$(\lambda(t - s))$ distribution.
2. For $0 \le s_1 < t_1 \le s_2 < t_2 \cdots \le s_k < t_k$ the variables $N(s_i, t_i); i = 1, \ldots, k$ are independent.

Approach 2:

Let $0 < S_1 < S_2 < \cdots$ be the times at which the particles arrive.

Let $T_i = S_i - S_{i-1}$ with $S_0 = 0$ by convention.

Then $T_1, T_2, \ldots$ are independent Exponential random variables with mean $1/\lambda$.

Note $P(T_i > x) = e^{-\lambda x}$ is called **survival function** of T_i.

Approaches are equivalent. Both are deductions of a model based on **local** behaviour of process.

Approach 3: Assume:

1. given all the points in $[0, t]$ the probability of 1 point in the interval $(t, t + h]$ is of the form
$$\lambda h + o(h)$$
2. given all the points in $[0, t]$ the probability of 2 or more points in interval $(t, t + h]$ is of the form
$$o(h)$$

All 3 approaches are equivalent. I show: 3 implies 1, 1 implies 2 and 2 implies 3. First explain o, O.

Notation: given functions f and g we write
$$f(h) = g(h) + o(h)$$
provided
$$\lim_{h \to 0} \frac{f(h) - g(h)}{h} = 0$$

[Aside: if there is a constant M such that
$$\limsup_{h \to 0} \left| \frac{f(h) - g(h)}{h} \right| \le M$$
we say
$$f(h) = g(h) + O(h)$$
Notation due to Landau. Another form is
$$f(h) = g(h) + O(h)$$
means there is $\delta > 0$ and M s.t. for all $|h| < \delta$
$$|f(h) - g(h)| \le Mh$$
Idea: $o(h)$ is tiny compared to h while $O(h)$ is (very) roughly the same size as h.]

Model 3 implies 1: Fix t, define $f_t(s)$ to be conditional probability of 0 points in $(t, t + s]$ given value of process on $[0, t]$.

Derive differential equation for f. Given process on $[0, t]$ and 0 points in $(t, t + s]$ probability of no points in $(t, t + s + h]$ is
$$f_{t+s}(h) = 1 - \lambda h + o(h)$$

Given the process on $[0,t]$ the probability of no points in $(t, t+s]$ is $f_t(s)$. Using $P(AB|C) = P(A|BC)P(B|C)$ gives

$$\begin{aligned} f_t(s+h) &= f_t(s)f_{t+s}(h) \\ &= f_t(s)(1 - \lambda h + o(h)) \end{aligned}$$

Now rearrange, divide by h to get

$$\frac{f_t(s+h) - f_t(s)}{h} = -\lambda f_t(s) + \frac{o(h)}{h}$$

Let $h \to 0$ and find

$$\frac{\partial f_t(s)}{\partial s} = -\lambda f_t(s)$$

Differential equation has solution

$$f_t(s) = f_t(0)\exp(-\lambda s) = \exp(-\lambda s)\,.$$

Notice: survival function of exponential rv..
General case. Notation: $N(t) = N(0,t)$.
$N(t)$ is a non-decreasing function of t. Let

$$P_k(t) = P(N(t) = k)$$

Evaluate $P_k(t+h)$ by conditioning on $N(s); 0 \le s < t$ and $N(t) = j$.
Given $N(t) = j$ probability that $N(t+h) = k$ is conditional probability of $k-j$ points in $(t, t+h]$.
So, for $j \le k-2$:

$$\begin{aligned} &P(N(t+h) = k|N(t) = j, N(s), 0 \le s < t) \\ &\qquad = o(h) \end{aligned}$$

For $j = k-1$ we have

$$\begin{aligned} &P(N(t+h) = k|N(t) = k-1, N(s), 0 \le s < t) \\ &\qquad = \lambda h + o(h) \end{aligned}$$

For $j = k$ we have

$$\begin{aligned} &P(N(t+h) = k|N(t) = k, N(s), 0 \le s < t) \\ &\qquad = 1 - \lambda h + o(h) \end{aligned}$$

N is increasing so only consider $j \le k$.

$$\begin{aligned} P_k(t+h) &= \sum_{j=0}^{k} P(N(t+h) = k|N(t) = j)P_j(t) \\ &= P_k(t)(1 - \lambda h) + \lambda h P_{k-1}(t) + o(h) \end{aligned}$$

Rearrange, divide by h and let $h \to 0$ t get

$$P_k'(t) = -\lambda P_k(t) + \lambda P_{k-1}(t)$$

For $k = 0$ the term P_{k-1} is dropped and

$$P_0'(t) = -\lambda P_0(t)$$

Using $P_0(0) = 1$ we get

$$P_0(t) = e^{-\lambda t}$$

Put this into the equation for $k = 1$ to get

$$P_1'(t) = -\lambda P_1(t) + \lambda e^{-\lambda t}$$

Multiply by $e^{\lambda t}$ to see

$$\left(e^{\lambda t} P_1(t)\right)' = \lambda$$

With $P_1(0) = 0$ we get

$$P_1(t) = \lambda t e^{-\lambda t}$$

For general k we have $P_k(0) = 0$ and

$$\left(e^{\lambda t} P_k(t)\right)' = \lambda e^{\lambda t} P_{k-1}(t)$$

Check by induction that

$$e^{\lambda t} P_k(t) = (\lambda t)^k / k!$$

Hence: $N(t)$ has Poisson(λt) distribution.

Similar ideas permit proof of

$$P(N(s,t) = k | N(u); 0 \le u \le s)$$

$$= \frac{\{\lambda(t-s)\}^k e^{-\lambda(t-s)}}{k!}$$

From which (by induction) we can prove that N has independent Poisson increments.

Exponential Interarrival Times

If N is a Poisson Process we define $T_1, T_2, \ldots$ to be the times between 0 and the first point, the first point and the second and so on.

Fact: $T_1, T_2, \ldots$ are iid exponential rvs with mean $1/\lambda$.

We already did T_1 rigorously. The event $T > t$ is exactly the event $N(t) = 0$. So

$$P(T > t) = \exp(-\lambda t)$$

which is the survival function of an exponential rv.

I do case of T_1, T_2. Let t_1, t_2 be two positive numbers and $s_1 = t_1$, $s_2 = t_1 + t_2$. Consider event

$$\{t_1 < T_1 \le t_1 + \delta_1\} \cap \{t_2 < T_2 \le t_2 + \delta_2\}.$$

This is almost the same as the intersection of the four events:

$$\begin{aligned} N(0, t_1] &= 0 \\ N(t_1, t_1 + \delta_1] &= 1 \\ N(t_1 + \delta_1, t_1 + \delta_1 + t_2] &= 0 \\ N(s_2 + \delta_1, s_2 + \delta_1 + \delta_2] &= 1 \end{aligned}$$

which has probability

$$e^{-\lambda t_1} \times \lambda \delta_1 e^{-\lambda \delta_1} \times e^{-\lambda t_2} \times \lambda \delta_2 e^{-\lambda \delta_2}$$

Divide by $\delta_1 \delta_2$ and let δ_1 and δ_2 go to 0 to get joint density of T_1, T_2 is

$$\lambda^2 e^{-\lambda t_1} e^{-\lambda t_2}$$

which is the joint density of two independent exponential variates.

More rigor:

- Find joint density of $S_1, \ldots, S_k$.

- Use **change of variables** to find joint density of $T_1,\ldots,T_k$.

First step: Compute

$$P(0 < S_1 \le s_1 < S_2 \le s_2 \cdots < S_k \le s_k)$$

This is just the event of exactly 1 point in each interval $(s_{i-1}, s_i]$ for $i = 1,\ldots,k-1$ ($s_0 = 0$) and at least one point in $(s_{k-1}, s_k]$ which has probability

$$\prod_1^{k-1}\left\{\lambda(s_i - s_{i-1})e^{-\lambda(s_i-s_{i-1})}\right\}\left(1-e^{-\lambda(s_k-s_{k-1})}\right)$$

Second step: write this in terms of joint cdf of $S_1,\ldots,S_k$. I do $k=2$:

$$P(0 < S_1 \le s_1 < S_2 \le s_2) = F_{S_1,S_2}(s_1,s_2) - F_{S_1,S_2}(s_1,s_1)$$

Notice tacit assumption $s_1 < s_2$.

Differentiate twice, that is, take

$$\frac{\partial^2}{\partial s_1 \partial s_2}$$

to get

$$f_{S_1,S_2}(s_1,s_2) = \frac{\partial^2}{\partial s_1 \partial s_2}\lambda s_1 e^{-\lambda s_1}\left(1-e^{-\lambda(s_2-s_1)}\right)$$

Simplify to

$$\lambda^2 e^{-\lambda s_2}$$

Recall tacit assumption to get

$$f_{S_1,S_2}(s_1,s_2) = \lambda^2 e^{-\lambda s_2} 1(0 < s_1 < s_2)$$

That completes the first part.

Now compute the joint cdf of T_1, T_2 by

$$F_{T_1,T_2}(t_1,t_2) = P(S_1 < t_1, S_2 - S_1 < t_2)$$

This is

$$\begin{aligned}
P(S_1 < t_1, &S_2 - S_1 < t_2) \\
&= \int_0^{t_1}\int_{s_1}^{s_1+t_2} \lambda^2 e^{-\lambda s_2} ds_2 ds_1 \\
&= \lambda \int_0^{t_1}\left(e^{-\lambda s_1} - e^{-\lambda(s_1+t_2)}\right) ds_1 \\
&= 1 - e^{-\lambda t_1} - e^{-\lambda t_2} + e^{-\lambda(t_1+t_2)}
\end{aligned}$$

Differentiate twice to get

$$f_{T_1,T_2}(t_1,t_2) = \lambda e^{-\lambda t_1}\lambda e^{-\lambda t_2}$$

which is the joint density of two independent exponential random variables.

Summary so far:

Have shown:

Instantaneous rates model implies independent Poisson increments model implies independent exponential interarrivals.

Next: show independent exponential interarrivals implies the instantaneous rates model.

Suppose $T_1, \ldots$ iid exponential rvs with means $1/\lambda$. Define N_t by $N_t = k$ if and only if

$$T_1 + \cdots + T_k \le t \le T_1 + \cdots + T_{k+1}$$

Let A be the event $N(s) = n(s); 0 < s \le t$. We are to show

$$P(N(t, t+h] = 1 | N(t) = k, A) = \lambda h + o(h)$$

and

$$P(N(t, t+h] \ge 2 | N(t) = k, A) = o(h)$$

If $n(s)$ is a possible trajectory consistent with $N(t) = k$ then n has jumps at points

$$s_1 \equiv t_1, s_2 \equiv t_1 + t_2, \ldots, s_k \equiv t_1 + \cdots + t_k < t$$

and at no other points in $(0, t]$.

So given $N(s) = n(s); 0 < s \le t$ with $n(t) = k$ we are essentially being given

$$T_1 = t_1, \ldots, T_k = t_k, T_{k+1} > t - s_k$$

and asked the conditional probabilty in the first case of the event B given by

$$t - s_k < T_{k+1} \le t - s_k + h < T_{k+2} + T_{k+1} .$$

Conditioning on $T_1, \ldots, T_k$ irrelevant (independence).

$$\begin{aligned} P(N(t, t+h] = 1 | N(t) = k, A)/h & \\ &= P(B | T_{k+1} > t - s_k)/h \\ &= \frac{P(B)}{he^{-\lambda(t - s_k)}} \end{aligned}$$

The numerator may be evaluated by integration:

$$P(B) = \int_{t-s_k}^{t-s_k+h} \int_{t-s_k+h-s_1}^{\infty} \lambda^2 e^{-\lambda(s_1+s_2)} ds_2 ds_1$$

Let $h \to 0$ to get the limit

$$P(N(t, t+h] = 1 | N(t) = k, A)/h \to \lambda$$

as required.

The computation of

$$\lim_{h \to 0} P(N(t, t+h] \ge 2 | N(t) = k, A)/h$$

is similar.

Properties of exponential rvs

Convolution: If X and Y independent rvs with densities f and g respectively and $Z = X + Y$ then

$$P(Z \le z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x) g(y) dy dx$$

Differentiating wrt z we get

$$f_Z(z) = \int_{-\infty}^{\infty} f(x) g(z - x) dx$$

This integral is called the **convolution** of densities f and g.

If $T_1,\ldots,T_n$ iid Exponential(λ) then $S_n = T_1+\cdots+T_n$ has a Gamma(n,λ) distribution. Density of S_n is

$$f_{S_n}(s) = \lambda(\lambda s)^{n-1}e^{-\lambda s}/(n-1)!$$

for $s > 0$.

Proof:

$$\begin{aligned} P(S_n > s) &= P(N(0,s] < n) \\ &= \sum_{j=0}^{n-1}(\lambda s)^j e^{-\lambda s}/j! \end{aligned}$$

Then

$$\begin{aligned} f_{S_n}(s) &= \frac{d}{ds}P(S_n \le s) \\ &= \frac{d}{ds}\left\{1 - P(S_n > s)\right\} \\ &= -\lambda \sum_{j=1} n-1 \left\{ j(\lambda s)^{j-1} - (\lambda s)^j \right\} \frac{e^{-\lambda s}}{j!} \\ &\quad + \lambda e^{-\lambda s} \\ &= \lambda e^{-\lambda s} \sum_{j=1}^{n-1} \left\{ \frac{(\lambda s)^j}{j!} - \frac{(\lambda s)^{j-1}}{(j-1)!} \right\} \\ &\quad + \lambda e^{-\lambda s} \end{aligned}$$

This telescopes to

$$f_{S_n}(s) = \lambda(\lambda s)^{n-1}e^{-\lambda s}/(n-1)!$$

Extreme Values: If $X_1,\ldots,X_n$ are independent exponential rvs with means $1/\lambda_1,\ldots,1/\lambda_n$ then $Y = \min\{X_1,\ldots,X_n\}$ has an exponential distribution with mean

$$\frac{1}{\lambda_1+\cdots+\lambda_n}$$

Proof:

$$\begin{aligned} P(Y > y) &= P(\forall k X_k > y) \\ &= \prod e^{-\lambda_k y} \\ &= e^{-\sum \lambda_k y} \end{aligned}$$

Memoryless Property: conditional distribution of $X - x$ given $X \ge x$ is exponential if X has an exponential distribution.

Proof:

$$\begin{aligned} &P(X - x > y | X \ge x) \\ &\qquad = \frac{P(X > x+y, X \ge x)}{P(X > x)} \\ &\qquad = \frac{PX > x+y)}{P(X \ge x)} \\ &\qquad = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} \\ &\qquad = e^{-\lambda y} \end{aligned}$$

Hazard Rates

The hazard rate, or instantaneous failure rate for a positive random variable T with density f and cdf F is

$$r(t) = \lim_{\delta \to 0} \frac{P(t < T \le t + \delta | T \ge t)}{\delta}$$

This is just

$$r(t) = \frac{f(t)}{1 - F(t)}$$

For an exponential random variable with mean $1/\lambda$ this is

$$h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

The exponential distribution has constant failure rate.

Weibull random variables have density

$$f(t|\lambda, \alpha) = \lambda(\lambda t)^{\alpha - 1} e^{-(\lambda t)^\alpha}$$

for $t > 0$. The corresponding survival function is

$$1 - F(t) = e^{-(\lambda t)^\alpha}$$

and the hazard rate is

$$r(t) = \lambda(\lambda t)^{\alpha - 1}$$

which is increasing for $\alpha > 1$, decreasing for $\alpha < 1$. For $\alpha = 1$ this is the exponential distribution.

Since

$$r(t) = \frac{dF(t)/dt}{1 - F(t)} = -\frac{d \log(1 - F(t))}{dt}$$

we can integrate to find

$$1 - F(t) = \exp\{-\int_0^t r(s) ds\}$$

so that r determines F and f.

Properties of Poisson Processes

1) If N_1 and N_2 are independent Poisson processes with rates λ_1 and λ_2, respectively, then $N = N_1 + N_2$ is a Poisson process with rate $\lambda_1 + \lambda_2$.

2) Suppose N is a Poisson process with rate λ. Suppose each point is marked with a label, say one of $L_1, \ldots, L_r$, independently of all other occurences. Suppose p_i is the probability that a given point receives label L_i. Let N_i count the points with label i (so that $N = N_1 + \cdots + N_r$). Then $N_1, \ldots, N_r$ are independent Poisson processes with rates $p_i \lambda$.

3) Suppose $U_1, U_2, \ldots$ independent rvs, each uniformly distributed on $[0, T]$. Suppose M is a Poisson(λT) random variable independent of the U's. Let

$$N(t) = \sum_1^M 1(U_i \le t)$$

Then N is a Poisson process on $[0, T]$ with rate λ.

4) Suppose N is a Poisson process with rate λ. Let $S_1 < S_2 < \cdots$ be the times at which points arrive Given $N(T) = n$ $S_1, \ldots, S_n$ have the same distribution as the order statistics of a sample of size n from the uniform distribution on $[0, T]$.

5) Given $S_{n+1} = T$, $S_1, \ldots, S_n$ have the same distribution as the order statistics of a sample of size n from the uniform distribution on $[0, T]$.

Indications of some proofs:

1) $N_1, \ldots, N_r$ independent Poisson processes rates λ_i, $N = \sum N_i$. Let A_h be the event of 2 or more points in N in the time interval $(t, t+h]$, B_h, the event of exactly one point in N in the time interval $(t, t+h]$.

Let A_{ih} and B_{ih} be the corresponding events for N_i.

Let H_t denote the history of the processes up to time t; we condition on H_t.

We are given:

$$P(A_{ih}|H_t) = o(h)$$

and

$$P(B_{ih}|H_t) = \lambda_i h + o(h) .$$

Note that

$$A_h \subset \bigcup_{i=1}^{r} A_{ih} \cup \bigcup_{i \neq j} (B_{ih} \cap B_{jh})$$

Since

$$\begin{aligned} P(B_{ih} \cap B_{jh}|H_t) &= P(B_{ih}|H_t)P(B_{jh}|H_t) \\ &= (\lambda_i h + o(h))(\lambda_j h + o(h)) \\ &= O(h^2) \\ &= o(h) \end{aligned}$$

we have checked one of the two infinitesimal conditions for a Poisson process.

Next let C_h be the event of no points in N in the time interval $(t, t+h]$ and C_{ih} the same for N_i. Then

$$\begin{aligned} P(C_h|H_t) &= P(\cap C_{ih}|H_t) \\ &= \prod P(C_{ih}|H_t) \\ &= \prod (1 - \lambda_i h + o(h)) \\ &= 1 - (\sum \lambda_i)h + o(h) \end{aligned}$$

shows

$$\begin{aligned} P(B_h|H_t) &= 1 - P(C_h|H_t) - P(A_h|H_t) \\ &= (\sum \lambda_i)h + o(h) \end{aligned}$$

Hence N is a Poisson process with rate $\sum \lambda_i$.

2) The infinitesimal approach used for 1 can do part of this. See text for rest. Events defined as in **1)**: The event B_{ih} that there is one point in N_i in $(t, t+h]$ is the event, B_h that there is exactly one point in any of the r processes together with a subset of A_h where there are two or more points in N in $(t, t+h]$ but exactly one is labeled i. Since $P(A_h|H_t) = o(h)$

$$\begin{aligned} P(B_{ih}|H_t) &= p_i P(B_h|H_t) + o(h) \\ &= p_i(\lambda h + o(h)) + o(h) \\ &= p_i \lambda h + o(h) \end{aligned}$$

Similarly, A_{ih} is a subset of A_h so

$$P(A_{ih}|H_t) = o(h)$$

This shows each N_i is Poisson with rate λp_i. To get independence requires more work; see the text for the algebraic method which is easier.

3) Fix $s < t$. Let $N(s,t)$ be the number of points in $(s,t]$. Given $N = n$ the conditional distribution of $N(s,t)$ is Binomial(n,p) with $p = (t-s)/T$. So

$$\begin{aligned}
P(N(s,t) = k)& \\
&= \sum_{n=k}^{\infty} P(N(s,t) = k, N = n) \\
&= \sum_{n=k}^{\infty} P(N(s,t) = k | N = n) P(N = n) \\
&= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{(\lambda T)^n}{n!} e^{-\lambda T} \\
&= \frac{e^{-\lambda T}}{k!} (\lambda T p)^k \sum_{n=k}^{\infty} \frac{(1-p)^{n-k} (\lambda T)^{n-k}}{(n-k)!} \\
&= \frac{e^{-\lambda T}}{k!} (\lambda T p)^k \sum_{m=0}^{\infty} (1-p)^m (\lambda T)^m / m! \\
&= \frac{e^{-\lambda T}}{k!} (\lambda T p)^k e^{\lambda T(1-p)} \\
&= \frac{e^{-\lambda(t-s)} (\lambda(t-s))^k}{k!}
\end{aligned}$$

4): Fix s_i, h_i for $i = 1, \ldots, n$ such that

$$0 < s_1 < s_1 + h_1 < s_2 < \cdots < s_n < s_n + h_n < T$$

Given $N(T) = n$ we compute the probability of the event

$$A = \bigcap_{i=1}^{n} \{s_i < S_i < s_i + h_i\}$$

Intersection of A, $N(T) = n$ is ($s_0 = h_0 = 0$):

$$\begin{aligned}
B \equiv& \\
&\bigcap_{i=1}^{n} \{N(s_{i-1} + h_{i-1}, s_i] = 0, N(s_i, s_i + h_i] = 1\} \\
&\qquad \cap \{N(s_n + h_n, T] = 0\}
\end{aligned}$$

whose probability is

$$\left(\prod \lambda h_i\right) e^{-\lambda T}$$

So

$$\begin{aligned}
P(A | N(t) = n) &= \frac{P(A, N(T) = n)}{P(N(T) = n)} \\
&= \frac{\lambda^n e^{-\lambda T} \prod h_i}{(\lambda T)^n e^{-\lambda T} / n!} \\
&= \frac{n! \prod h_i}{T^n}
\end{aligned}$$

Divide by $\prod h_i$ and let all h_i go to 0 to get joint density of $S_1, \ldots, S_n$ is

$$\frac{n!}{T^n} 1(0 < s_1 < \cdots < s_n < T)$$

which is the density of order statistics from a Uniform$[0, T]$ sample of size n.

5) Replace the event $S_{n+1} = T$ with $T < S_{n+1} < T + h$. With A as before we want

$$P(A|T < S_{n+1} < T + h) = \frac{P(B, N(T, T+h] \geq 1)}{P(T < S_{n+1} < T + h)}$$

Note that B is independent of $\{N(T, T+h] \geq 1\}$ and that we have already found the limit

$$\frac{P(B)}{\prod h_i} \to \lambda^n e^{-\lambda T}$$

We are left to compute the limit of

$$\frac{P(N(T, T+h] \geq 1)}{P(T < S_{n+1} < T + h)}$$

The denominator is

$$\sum_{k=0}^{n} P(N(0, T] = k, N(T, T+h] = n + 1 - k) + o(h) = P(N(0, T] = n)\lambda h + o(h)$$

Thus

$$\frac{P(N(T, T+h] \geq 1)}{P(T < S_{n+1} < T + h)} = \frac{\lambda h + o(h)}{\frac{(\lambda T)^n}{n!} e^{-\lambda T} \lambda h + o(h)} \to \frac{n!}{(\lambda T)^n e^{-\lambda T}}$$

This gives the conditional density of $S_1, \ldots, S_n$ given $S_{n+1} = T$ as in **4)**.

Inhomogeneous Poisson Processes

The idea of hazard rate can be used to extend the notion of Poisson Process. Suppose $\lambda(t) \geq 0$ is a function of t. Suppose N is a counting process such that

$$P(N(t + h) = k + 1 | N(t) = k, H_t) = \lambda(t)h + o(h)$$

and

$$P(N(t + h) \geq k + 2 | N(t) = k, H_t) = o(h)$$

Then N has independent increments and $N(t + s) - N(t)$ has a Poisson distribution with mean

$$\int_t^{t+s} \lambda(u) du$$

If we put

$$\Lambda(t) = \int_0^t \lambda(u) du$$

then mean of $N(t + s) - N(t)$ is $\Lambda(t + s) - \Lambda(t)$.

Jargon: λ is the **intensity** or **instaneous intensity** and Λ the **cumulative intensity**.

Can use the model with Λ any non-decreasing right continuous function, possibly without a derivative. This allows ties.

Compound Poisson Processes

Imagine insurance claims arise at times of a Poisson process, $N(t)$, (more likely for an inhomogeneous process).

Let Y_i be the value of the ith claim associated with the point whose time is S_i.

Assume that the Y's are independent of each other and of N.

Let

$$\mu = \mathrm{E}(Y_i) \text{ and } \sigma^2 = \mathrm{var}(Y_i)$$

Let

$$X(t) = \sum_{i=1}^{N(t)} Y_i$$

be the total claim up to time t. We call X a compound Poisson Process.

Useful properties:

$$\begin{aligned}
\mathrm{E}\{X(t)|N(t)\} &= N(t)\mu \\
\mathrm{Var}\{X(t)|N(t)\} &= N(t)\sigma^2 \\
\mathrm{E}\{X(t)\} &= \mu \mathrm{E}\{N(t)\} \\
&= \mu\lambda t \\
\mathrm{Var}\{X(t)\} &= \mathrm{Var}\left[\mathrm{E}\{X(t)|N(t)\}\right] \\
&\quad + \mathrm{E}\left[\mathrm{Var}\{X(t)|N(t)\}\right] \\
&= \lambda t \mu^2 + \lambda t \sigma^2
\end{aligned}$$

(Look at all familiar? See homework.)

Continuous Time Markov Chains

Consider a population of single celled organisms in a stable environment.

Fix short time interval, length h.

Each organism has some probability of dividing to produce two organisms and some other probability of dying.

We might suppose:

- Different organisms behave independently.
- Probability of division is λh plus $o(h)$.
- Probability of death is μh plus $o(h)$.
- Probability that an organism divides twice (or divides once and dies) in the interval of length h is $o(h)$.

Tacit assumptions:

Constants of proportionality do not depend on time: "stable environment".

Constants do not depend on organism: organisms are all similar and live in similar environments.

$Y(t)$: total population at time t.

$\mathcal{H}_t$: history of process up to time t.

Condition on event $Y(t) = n$.

Probability of two or more divisions (more than one division by a single organism or two or more organisms dividing) is $o(h)$.

Probability of both a division and a death or of two or more deaths is $o(h)$.

So probability of exactly 1 division by any one of the n organisms is $n\lambda h + o(h)$.

Similarly probability of 1 death is $n\mu h + o(h)$.

We deduce:

$$\begin{aligned}
P(Y(t+h) &= n+1|Y(t)=n,\mathcal{H}_t) \\
&= n\lambda h + o(h) \\
P(Y(t+h) &= n-1|Y(t)=n,\mathcal{H}_t) \\
&= n\mu h + o(h) \\
P(Y(t+h) &= n|Y(t)=n,\mathcal{H}_t) \\
&= 1 - n(\lambda+\mu)h + o(h) \\
P(Y(t+h) &\notin \{n-1,n,n+1\}|Y(t)=n,\mathcal{H}_t) \\
&= o(h)
\end{aligned}$$

These equations lead to:

$$\begin{aligned}
P(Y(t+s) = j|Y(s) = i,\mathcal{H}_s) & \\
&= P(Y(t+s)=j|Y(s)=i) \\
&= P(Y(t)=j|Y(0)=i)
\end{aligned}$$

This is the **Markov Property**.

Definition:A process $\{Y(t); t \geq 0\}$ taking values in S, a finite or countable state space is a Markov Chain if

$$\begin{aligned}
P(Y(t+s) = j|Y(s) = i,\mathcal{H}_s) & \\
&= P(Y(t+s)=j|Y(s)=i) \\
&\equiv \mathbf{P}_{ij}(s,s+t)
\end{aligned}$$

Definition:A Markov chain Y has **stationary transitions** if

$$\mathbf{P}_{ij}(s,s+t) = \mathbf{P}_{ij}(0,t) \equiv \mathbf{P}_{ij}(t)$$

From now on: our chains have stationary transitions.

Summary of Markov Process Results

- Chapman-Kolmogorov equations:

$$\mathbf{P}_{ik}(t+s) = \sum_j \mathbf{P}_{ij}(t)\mathbf{P}_{jk}(s)$$

- Exponential holding times: starting from state i time, T_i, until process leaves i has exponential distribution, rate denoted v_i.
- Sequence of states visited, $Y_0, Y_1, Y_2, \ldots$ is Markov chain – transition matrix has $\mathbf{P}_{ii} = 0$. Y sometimes called **skeleton**.
- **Communicating classes** defined for skeleton chain. Usually assume chain has 1 communicating class.
- Periodicity irrelevant because of continuity of exponential distribution.
- Instantaneous transition rates from i to j:

$$q_{ij} = v_i \mathbf{P}_{ij}$$

- Kolmogorov backward equations:

$$\mathbf{P}'_{ij}(t) = \sum_{k \neq i} q_{ik}\mathbf{P}_{kj}(t) - v_i \mathbf{P}_{ij}(t)$$

- Kolmogorov forward equations:

$$\mathbf{P}'_{ij}(t) = \sum_{k \neq j} q_{kj} \mathbf{P}_{ik}(t) - v_i \mathbf{P}_{ij}(t)$$

- For strongly recurrent chains with a single communicating class:

$$\mathbf{P}_{ij}(t) \to \pi_j$$

- Stationary initial probabilities π_i satisfy:

$$v_j \pi_j = \sum_{k \neq j} q_{kj} \pi_k$$

- Transition probabilities given by

$$\mathbf{P}(t) = e^{\mathbf{R}t}$$

  where $\mathbf{R}$ has entries

$$\mathbf{R}_{ij} = \begin{cases} q_{ij} & i \neq j \\ -v_i & i = j \end{cases}$$

- Process is a **Birth and Death** process if

$$\mathbf{P}_{ij} = 0 \text{ if } |i - j| > 1$$

  In this case we write λ_i for the instantaneous "birth" rate:

$$P(X(t+h) = i + 1 | X_t = i) = \lambda_i h + o(h)$$

  and μ_i for the instantaneous "death" rate:

$$P(X(t+h) = i - 1 | X_t = i) = \mu_i h + o(h)$$

  We have

$$q_{ij} = \begin{cases} 0 & |i - j| > 1 \\ \lambda_i & j = i + 1 \\ \mu_i & j = i - 1 \end{cases}$$

- If all $\mu_i = 0$ then process is a **pure birth** process. If all $\lambda_i = 0$ a **pure death** process.
- Birth and Death process have stationary distribution

$$\pi_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n}\right)}$$

  Necessary condition for existence of π is

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} < \infty$$

- **Linear** birth and death processes have

$$\lambda_n = (n+1)\lambda \qquad \mu_n = n\mu$$

  In this case (note *immigration*)

$$\pi_n = \frac{(\lambda/\mu)^n}{1 - (\lambda/\mu)}$$

  provided $\lambda < \mu$.

Detailed development

Suppose X a Markov Chain with stationary transitions. Then

$$\begin{aligned}
&P(X(t+s)=k|X(0)=i)\\
&\quad=\sum_j P(X(t+s)=k,X(t)=j|X(0)=i)\\
&\quad=\sum_j P(X(t+s)=k|X(t)=j,X(0)=i)\\
&\qquad\times P(X(t)=j|X(0)=i)\\
&\quad=\sum_j P(X(t+s)=k|X(t)=j)\\
&\qquad\times P(X(t)=j|X(0)=i)\\
&\quad=\sum_j P(X(s)=k|X(0)=j)\\
&\qquad\times P(X(t)=j|X(0)=i)
\end{aligned}$$

This shows

$$\mathbf{P}(t+s)=\mathbf{P}(t)\mathbf{P}(s)$$

which is the Chapman-Kolmogorov equation.

Now consider the chain starting from i and let T_i be the first t for which $X(t)\neq i$. Then T_i is a stopping time.

[Technically:

$$\{T_i\le t\}\in\mathcal{H}_t$$

for each t.] Then

$$\begin{aligned}
&P(T_i>t+s|T_i>s,X(0)=i)\\
&\quad=P(T_i>t+s|X(u)=i;0\le u\le s)\\
&\quad=P(T_i>t|X(0)=i)
\end{aligned}$$

by the Markov property.

Conclusion: given $X(0)=i$, T_i has memoryless property so T_i has an exponential distribution. Let v_i be the rate parameter.

Embedded Chain: Skeleton

Let $T_1<T_2<\cdots$ be the stopping times at which transitions occur. Then $X_n=X(T_n)$. The sequence X_n is a Markov chain by the Markov property. That $\mathbf{P}_{ii}=0$ reflects the fact that $P(X(T_{n+1})=X(T_n))=0$ by design.

As before we say $i\rightsquigarrow j$ if $\mathbf{P}_{ij}(t)>0$ for some t. It is fairly clear that $i\rightsquigarrow j$ for the $X(t)$ if and only if $i\rightsquigarrow j$ for the embedded chain X_n.

We say $i\leftrightarrow j$ (i and j communicate) if $i\rightsquigarrow j$ and $j\rightsquigarrow i$.

Now consider

$$P(X(t+h)=j|X(t)=i,\mathcal{H}_t)$$

Suppose the chain has made n transitions so far so that $T_n<t<T_{n+1}$. Then the event $X(t+h)=j$ is, except for possibilities of probability $o(h)$ the event that

$$t<T_{n+1}\le t+h \text{ and } X_{n+1}=j$$

The probability of this is

$$(v_ih+o(h))\mathbf{P}_{ij}=v_i\mathbf{P}_{ij}h+o(h)$$

Kolmogorov's Equations

The Chapman-Kolmogorov equations are

$$\mathbf{P}(t+h) = \mathbf{P}(t)\mathbf{P}(h)$$

Subtract $\mathbf{P}(t)$ from both sides, divide by h and let $h \to 0$. Remember that $\mathbf{P}(0)$ is the identity. We find

$$\frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} = \frac{\mathbf{P}(t)(\mathbf{P}(h) - \mathbf{P}(0))}{h}$$

which gives

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{P}'(0)$$

The Chapman-Kolmogorov equations can also be written

$$\mathbf{P}(t+h) = \mathbf{P}(h)\mathbf{P}(t)$$

Now subtracting $\mathbf{P}(t)$ from both sides, dividing by h and letting $h \to 0$ gives

$$\mathbf{P}'(t) = \mathbf{P}'(0)\mathbf{P}(t)$$

Look at these equations in component form:

$$\mathbf{P}'(t) = \mathbf{P}'(0)\mathbf{P}(t)$$

becomes

$$\mathbf{P}'_{ij}(t) = \sum_k \mathbf{P}'_{ik}(0)\mathbf{P}_{kj}(t)$$

For $i \neq k$ our calculations of instantaneous transition rates gives

$$\mathbf{P}'_{ik}(0) = v_i \mathbf{P}_{ik}$$

For $i = k$ we have

$$P(X(h) = i | X(0) = i) = e^{-v_i h} + o(h)$$

($X(h) = i$ either means $T_i > h$ which has probability $e^{-v_i h}$ or there have been two or more transitions in $[0, h]$, a possibility of probability $o(h)$.) Thus

$$\mathbf{P}'_{ii}(0) = -v_i$$

Let $\mathbf{R}$ be the matrix with entries

$$\mathbf{R}_{ij} = \begin{cases} q_{ij} \equiv v_i \mathbf{P}_{ij} & i \neq j \\ -v_i & i = j \end{cases}$$

That is

$$\mathbf{R} = \mathbf{P}'(0).$$

$\mathbf{R}$ is the **infinitesimal generator** of the chain.
Thus

$$\mathbf{P}'(t) = \mathbf{P}'(0)\mathbf{P}(t)$$

becomes

$$\begin{aligned} \mathbf{P}'_{ij}(t) &= \sum_k \mathbf{R}_{ik}\mathbf{P}_{kj}(t) \\ &= \sum_{k \neq i} q_{ik}\mathbf{P}_{kj}(t) - v_i \mathbf{P}_{ij}(t) \end{aligned}$$

Called **Kolmogorov's backward equations**.

On the other hand

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{P}'(0)$$

becomes

$$\begin{aligned}\mathbf{P}'_{ij}(t) &= \sum_k \mathbf{P}_{ik}(t)\mathbf{R}_{kj} \\ &= \sum_{k\neq j} q_{kj}\mathbf{P}_{ik}(t) - v_j\mathbf{P}_{ij}(t)\end{aligned}$$

These are **Kolmogorov's forward equations**.

Remark: When the state space is infinite the forward equations may not be justified. In deriving them we interchanged a limit with an infinite sum; the interchange is always justified for the backward equations but not for forward.

Example: $S = \{0, 1\}$. Then

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and the chain is otherwise specified by v_0 and v_1. The matrix $\mathbf{R}$ is

$$\mathbf{R} = \begin{bmatrix} -v_0 & v_0 \\ v_1 & -v_1 \end{bmatrix}$$

The backward equations become

$$\begin{aligned}\mathbf{P}'_{00}(t) &= v_0\mathbf{P}_{10}(t) - v_0\mathbf{P}_{00}(t) \\ \mathbf{P}'_{01}(t) &= v_0\mathbf{P}_{11}(t) - v_0\mathbf{P}_{01}(t) \\ \mathbf{P}'_{10}(t) &= v_1\mathbf{P}_{00}(t) - v_1\mathbf{P}_{10}(t) \\ \mathbf{P}'_{11}(t) &= v_1\mathbf{P}_{01}(t) - v_1\mathbf{P}_{11}(t)\end{aligned}$$

while the forward equations are

$$\begin{aligned}\mathbf{P}'_{00}(t) &= v_1\mathbf{P}_{01}(t) - v_0\mathbf{P}_{00}(t) \\ \mathbf{P}'_{01}(t) &= v_0\mathbf{P}_{00}(t) - v_1\mathbf{P}_{01}(t) \\ \mathbf{P}'_{10}(t) &= v_1\mathbf{P}_{11}(t) - v_0\mathbf{P}_{10}(t) \\ \mathbf{P}'_{11}(t) &= v_0\mathbf{P}_{10}(t) - v_1\mathbf{P}_{11}(t)\end{aligned}$$

Add $v_1\times$first plus $v_0\times$third backward equations to get

$$v_1\mathbf{P}'_{00}(t) + v_0\mathbf{P}'_{10}(t) = 0$$

so

$$v_1\mathbf{P}_{00}(t) + v_0\mathbf{P}_{10}(t) = c$$

Put $t = 0$ to get $c = v_1$. This gives

$$\mathbf{P}_{10}(t) = \frac{v_1}{v_0}\{1 - \mathbf{P}_{00}(t)\}$$

Plug this back in to the first equation and get

$$\mathbf{P}'_{00}(t) = v_1 - (v_1 + v_0)\mathbf{P}_{00}(t)$$

Multiply by $e^{(v_1+v_0)t}$ and get

$$\left\{e^{(v_1+v_0)t}\mathbf{P}_{00}(t)\right\}' = v_1 e^{(v_1+v_0)t}$$

which can be integrated to get

$$\mathbf{P}_{00}(t) = \frac{v_1}{v_0+v_1} + \frac{v_0}{v_0+v_1} e^{-(v_1+v_0)t}$$

Alternative calculation:

$$\mathbf{R} = \begin{bmatrix} -v_0 & v_0 \\ v_1 & -v_1 \end{bmatrix}$$

can be written as

$$\mathbf{M}\mathbf{\Lambda}\mathbf{M}^{-1}$$

where

$$\mathbf{M} = \begin{bmatrix} 1 & v_0 \\ \\ 1 & -v_1 \end{bmatrix}$$

$$\mathbf{M}^{-1} = \begin{bmatrix} \frac{v_1}{v_0+v_1} & \frac{v_0}{v_0+v_1} \\ \\ \frac{1}{v_0+v_1} & \frac{-1}{v_0+v_1} \end{bmatrix}$$

and

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & 0 \\ \\ 0 & -(v_0+v_1) \end{bmatrix}$$

Then

$$\begin{aligned} e^{\mathbf{R}t} &= \sum_0^\infty \mathbf{R}^n t^n/n! \\ &= \sum_0^\infty \left(\mathbf{M}\mathbf{\Lambda}\mathbf{M}^{-1}\right)^n \frac{t^n}{n!} \\ &= \mathbf{M}\left(\sum_0^\infty \mathbf{\Lambda}^n \frac{t^n}{n!}\right)\mathbf{M}^{-1} \end{aligned}$$

Now

$$\sum_0^\infty \mathbf{\Lambda}^n \frac{t^n}{n!} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-(v_0+v_1)t} \end{bmatrix}$$

so we get

$$\begin{aligned} \mathbf{P}(t) &= e^{\mathbf{R}t} \\ &= \mathbf{M}\begin{bmatrix} 1 & 0 \\ 0 & e^{-(v_0+v_1)t} \end{bmatrix}\mathbf{M}^{-1} \\ &= \mathbf{P}^\infty - \frac{e^{-(v_0+v_1)t}}{v_0+v_1}\mathbf{R} \end{aligned}$$

where

$$\mathbf{P}^\infty = \begin{bmatrix} \frac{v_1}{v_0+v_1} & \frac{v_0}{v_0+v_1} \\ \\ \frac{v_1}{v_0+v_1} & \frac{v_0}{v_0+v_1} \end{bmatrix}$$

Notice: rows of $\mathbf{P}^\infty$ are a stationary initial distribution. If rows are π then

$$\mathbf{P}^\infty = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \pi \equiv \mathbf{1}\pi$$

so

$$\pi \mathbf{P}^\infty = (\pi \mathbf{1})\pi = \pi$$

Moreover

$$\pi \mathbf{R} = \mathbf{0}$$

Fact: $\pi_0 = v_1/(v_0+v_1)$ is long run fraction of time in state 0.
Fact:

$$\frac{1}{T}\int_0^T f(X(t))dt \to \sum_j \pi_j f(j)$$

Ergodic Theorem in continuous time.

Birth and Death Processes

Consider a population of $X(t)$ individuals. Suppose in next time interval $(t, t+h)$ probability of population increase of 1 (called a birth) is $\lambda_i h + o(h)$ and probability of decrease of 1 (death) is $\mu_i h + o(h)$.

Jargon: X is a birth and death process.

Special cases:

All $\mu_i = 0$; called a **pure birth** process.

All $\lambda_i = 0$ (0 is absorbing): **pure death** process.

$\lambda_n = n\lambda$ and $\mu_n = n\mu$ is a **linear** birth and death process.

$\lambda_n \equiv 1$, $\mu_n \equiv 0$: Poisson Process.

$\lambda_n = n\lambda + \theta$ and $\mu_n = n\mu$ is a **linear** birth and death process with immigration.

Queuing Theory

Ingredients of Queuing Problem:

1: Queue input process.

2: Number of servers

3: Queue discipline: first come first serve? last in first out? pre-emptive priorities?

4: Service time distribution.

Example:Imagine customers arriving at a facility at times of a Poisson Process N with rate λ. This is the input process, denoted M (for Markov) in queuing literature.

Single server case:

Service distribution: exponential service times, rate μ.

Queue discipline: first come first serve.

$X(t)$ = number of customers in line at time t.

X is a Markov process called $M/M/1$ queue:

$$v_i = \lambda + \mu$$

$$\mathbf{P}_{ij} = \begin{cases} \frac{\mu}{\mu+\lambda} & j = i-1 \ge 0 \\ \frac{\lambda}{\mu+\lambda} & j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Example:$M/M/\infty$ queue:

Customers arrive according to PP rate λ. Each customer begins service immediately. $X(t)$ is number being served at time t. X is a birth and death process with

$$v_n = \lambda + n\mu$$

and

$$\mathbf{P}_{ij} = \begin{cases} \frac{i\mu}{i\mu+\lambda} & j = i-1 \ge 0 \\ \frac{\lambda}{i\mu+\lambda} & j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

Stationary Initial Distributions

We have seen that a stationary initial distribution π is a probability vector solving

$$\pi\mathbf{R} = \mathbf{0}$$

Rewrite this as

$$v_j\pi_j = \sum_{i\neq j} q_{ij}\pi_i$$

Interpretation: LHS is rate at which process leaves state j; process is in state j a fraction π_j of time and then makes transition at rate v_j. RHS is total rate of arrival in state j. For each state $i \neq j$ π_i is fraction of time spent in state i and then q_{ij} the instantaneous rate of transition from i to j.

So equation says:

Rate of departure from j balances rate of arrival to j. This is called balance.

Application to birth and death processes:

Equation is

$$(\lambda_j + \mu_j)\pi_j = \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1}$$

for $j \geq 1$ and

$$\lambda_0\pi_0 = \mu_1\pi_1$$

Notice that this permits the recursion

$$\begin{aligned}
\pi_1 &= \frac{\lambda_0}{\mu_1}\pi_0 \\
\pi_2 &= \frac{\lambda_1+\mu_1}{\mu_2}\pi_1 - \frac{\lambda_0}{\mu_2}\pi_0 \\
&= \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0
\end{aligned}$$

which extends by induction to

$$\pi_k = \frac{\lambda_0\cdots\lambda_{k-1}}{\mu_1\cdots\mu_k}\pi_0$$

Apply $\sum \pi_k = 1$ to get

$$\pi_0\left(1+\sum_{n=1}^\infty \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}\right) = 1$$

This gives the formula announced:

$$\pi_n = \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n\left(1+\sum_{n=1}^\infty \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}\right)}$$

If

$$\sum_{n=1}^\infty \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n} < \infty$$

then we have defined a probability vector which solves

$$\pi\mathbf{R} = \mathbf{0}$$

Since

$$\mathbf{P}' = \mathbf{R}\mathbf{P}$$

we see that

$$\{\pi\mathbf{P}(t)\}' = 0$$

so that $\pi\mathbf{P}(t)$ is constant. Put $t = 0$ to discover that the constant is π.

Brownian Motion

For fair random walk Y_n = number of heads minus number of tails,

$$Y_n = U_1 + \cdots + U_n$$

where the U_i are independent and

$$P(U_i = 1) = P(U_i = -1) = \frac{1}{2}$$

Notice:

$$\mathrm{E}(U_i) = 0$$
$$\mathrm{Var}(U_i) = 1$$

Recall central limit theorem:

$$\frac{U_1 + \cdots + U_n}{\sqrt{n}} \Rightarrow N(0, 1)$$

Now: rescale time axis so that n steps take 1 time unit and vertical axis so step size is $1/\sqrt{n}$.

We now turn these pictures into a stochastic process:

For $\frac{k}{n} \le t < \frac{k+1}{n}$ we define

$$X_n(t) = \frac{U_1 + \cdots + U_k}{\sqrt{n}}$$

Notice:

$$\mathrm{E}(X_n(t)) = 0$$

and

$$\mathrm{Var}(X_n(t)) = \frac{k}{n}$$

As $n \to \infty$ with t fixed we see $k/n \to t$. Moreover:

$$\frac{U_1 + \cdots + U_k}{\sqrt{k}} = \sqrt{\frac{n}{k}} X_n(t)$$

converges to $N(0, 1)$ by the central limit theorem. Thus

$$X_n(t) \Rightarrow N(0, t)$$

Another observation: $X_n(t + s) - X_n(t)$ is independent of $X_n(t)$ because the two rvs involve sums of different U_i.

Conclusions.

As $n \to \infty$ the processes X_n converge to a process X with the properties:

1. $X(t)$ has a $N(0, t)$ distribution.

2. X has independent increments: if

$$0 = t_0 < t_1 < t_2 < \cdots < t_k$$

then

$$X(t_1) - X(t_0), \ldots, X(t_k) - X(t_{k-1})$$

are independent .

3. The increments are **stationary**:

$$X(t + s) - X(s) \sim N(0, t)$$

regardless of s.

4. $X(0) = 0$.

Definition:Any process satisfying 1-4 above is a Brownian motion.

Properties of Brownian motion

- Suppose $t > s$. Then

$$\begin{aligned}
\mathrm{E}(X(t)|X(s)) &= \mathrm{E}\left\{X(t) - X(s) + X(s)|X(s)\right\} \\
&= \mathrm{E}\left\{X(t) - X(s)|X(s)\right\} \\
&\quad + \mathrm{E}\left\{X(s)|X(s)\right\} \\
&= 0 + X(s) = X(s)
\end{aligned}$$

Notice the use of independent increments and of $\mathrm{E}(Y|Y) = Y$.

- Again if $t > s$:

$$\begin{aligned}
&\mathrm{Var}\left\{X(t)|X(s)\right\} \\
&\quad = \mathrm{Var}\left\{X(t) - X(s) + X(s)|X(s)\right\} \\
&\quad = \mathrm{Var}\left\{X(t) - X(s)|X(s)\right\} \\
&\quad = \mathrm{Var}\left\{X(t) - X(s)\right\} \\
&\quad = t - s
\end{aligned}$$

Suppose $t < s$. Then $X(s) = X(t) + \{X(t) - X(s)\}$ is a sum of two independent normal variables. Do following calculation:

$X \sim N(0, \sigma^2)$, and $Y \sim N(0, \tau^2)$ independent. $Z = X + Y$.

Compute conditional distribution of X given Z:

$$\begin{aligned}
f_{X|Z}(x|z) &= \frac{f_{X,Z}(x, z)}{f_Z(z)} \\
&= \frac{f_{X,Y}(x, z - x)}{f_Z(z)} \\
&= \frac{f_X(x) f_Y(z - x)}{f_Z(z)}
\end{aligned}$$

Now Z is $N(0, \gamma^2)$ where $\gamma^2 = \sigma^2 + \tau^2$ so

$$\begin{aligned}
f_{X|Z}(x|z) &= \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} \frac{1}{\tau\sqrt{2\pi}} e^{-(z-x)^2/(2\tau^2)}}{\frac{1}{\gamma\sqrt{2\pi}} e^{-z^2/(2\gamma^2)}} \\
&= \frac{\gamma}{\tau\sigma\sqrt{2\pi}} \exp\{-(x - a)^2/(2b^2)\}
\end{aligned}$$

for suitable choices of a and b. To find them compare coefficients of x^2, x and 1.

Coefficient of x^2:

$$\frac{1}{b^2} = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$$

so $b = \tau\sigma/\gamma$.

Coefficient of x:

$$\frac{a}{b^2} = \frac{z}{\tau^2}$$

so that

$$a = b^2 z/\tau^2 = \frac{\sigma^2}{\sigma^2 + \tau^2} z$$

Finally you should check that

$$\frac{a^2}{b^2} = \frac{z^2}{\tau^2} - \frac{z^2}{\gamma^2}$$

to make sure the coefficients of 1 work out as well.

Conclusion: given $Z = z$ the conditional distribution of X is $N(a, b^2)$ with a and b as above.

Application to Brownian motion:

- For $t < s$ let X be $X(t)$ and Y be $X(s) - X(t)$ so $Z = X + Y = X(s)$. Then $\sigma^2 = t$, $\tau^2 = s - t$ and $\gamma^2 = s$. Thus

  $$b^2 = \frac{(s-t)t}{s}$$

  and

  $$a = \frac{t}{s}X(s)$$

  SO:

  $$\mathrm{E}(X(t)|X(s)) = \frac{t}{s}X(s)$$

  and

  $$\mathrm{Var}(X(t)|X(s)) = \frac{(s-t)t}{s}$$

The Reflection Principle

Tossing a fair coin:

HTHHHTHTHHTHHHTTHTH	5 more heads than tails
THTTTHTHTTHTTTHHTHT	5 more tails than heads

Both sequences have the same probability.

So: for random walk starting at stopping time:

Any sequence with k more heads than tails in next m tosses is matched to sequence with k more tails than heads. Both sequences have same prob.

Suppose Y_n is a fair ($p = 1/2$) random walk. Define

$$M_n = \max\{Y_k, 0 \le k \le n\}$$

Compute $P(M_n \ge x)$? Trick: Compute

$$P(M_n \ge x, Y_n = y)$$

First: if $y \ge x$ then

$$\{M_n \ge x, Y_n = y\} = \{Y_n = y\}$$

Second: if $M_n \ge x$ then

$$T \equiv \min\{k : Y_k = x\} \le n$$

Fix $y < x$. Consider a sequence of H's and T's which leads to say $T = k$ and $Y_n = y$.

Switch the results of tosses $k+1$ to n to get a sequence of H's and T's which has $T = k$ and $Y_n = x + (x - y) = 2x - y > x$. This proves

$$P(T = k, Y_n = y) = P(T = k, Y_n = 2x - y)$$

This is true for each k so

$$\begin{aligned} P(M_n \ge x, Y_n = y) &= P(M_n \ge x, Y_n = 2x - y) \\ &= P(Y_n = 2x - y) \end{aligned}$$

Finally, sum over all y to get

$$P(M_n \geq x) = \sum_{y \geq x} P(Y_n = y) \\ + \sum_{y < x} P(Y_n = 2x - y)$$

Make the substitution $k = 2x - y$ in the second sum to get

$$\begin{aligned} P(M_n \geq x) &= \sum_{y \geq x} P(Y_n = y) \\ &\quad + \sum_{k > x} P(Y_n = k) \\ &= 2 \sum_{k > x} P(Y_n = k) + P(Y_n = x) \end{aligned}$$

Brownian motion version:

$$M_t = \max\{X(s); 0 \leq s \leq t\}$$

$$T_x = \min\{s : X(s) = x\}$$

(called hitting time for level x). Then

$$\{T_x \leq t\} = \{M_t \geq x\}$$

Any path with $T_x = s < t$ and $X(t) = y < x$ is matched to an equally likely path with $T_x = s < t$ and $X(t) = 2x - y > x$.

So for $y > x$

$$P(M_t \geq x, X(t) > y) = P(X(t) > y)$$

while for $y < x$

$$P(M_t \geq x, X(t) < y) = P(X(t) > 2x - y)$$

Let $y \to x$ to get

$$\begin{aligned} P(M_t \geq x, X(t) > x) &= P(M_t \geq x, X(t) < x) \\ &= P(X(t) > x) \end{aligned}$$

Adding these together gives

$$\begin{aligned} P(M_t > x) &= 2P(X(t) > x) \\ &= 2P(N(0,1) > x/\sqrt{t}) \end{aligned}$$

Hence M_t has the distribution of $|N(0,t)|$.

On the other hand in view of

$$\{T_x \leq t\} = \{M_t \geq x\}$$

the density of T_x is

$$\frac{d}{dt} 2P(N(0,1) > x/\sqrt{t})$$

Use the chain rule to compute this. First

$$\frac{d}{dy} P(N(0,1) > y) = -\phi(y)$$

where ϕ is the standard normal density

$$\phi(y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}}$$

because $P(N(0,1) > y)$ is 1 minus the standard normal cdf.

So

$$\begin{aligned}
&\frac{d}{dt} 2P(N(0,1) > x/\sqrt{t}) \\
&\qquad = -2\phi(x/\sqrt{t}) \frac{d}{dt}(x/\sqrt{t}) \\
&\qquad = \frac{x}{\sqrt{2\pi} t^{3/2}} \exp\{-x^2/(2t)\}
\end{aligned}$$

This density is called the **Inverse Gaussian** density. T_x is called a **first passage time**

NOTE: the preceding is a density when viewed as a function of the variable t.

Martingales

A stochastic process $M(t)$ indexed by either a discrete or continuous time parameter t is a **martingale** if:

$$\mathrm{E}\{M(t)|M(u); 0 \le u \le s\} = M(s)$$

whenever $s < t$.

Examples

- A fair random walk is a martingale.
- If $N(t)$ is a Poisson Process with rate λ then $N(t) - \lambda t$ is a martingale.
- Standard Brownian motion (defined above) is a martingale.

Note: Brownian motion with drift is a process of the form

$$X(t) = \sigma B(t) + \mu t$$

where B is **standard** Brownian motion, introduced earlier. X is a martingale if $\mu = 0$. We call μ the **drift**

- If $X(t)$ is a Brownian motion with drift then

  $$Y(t) = e^{X(t)}$$

  is a geometric Brownian motion. For suitable μ and σ we can make $Y(t)$ a martingale.
- If a gambler makes a sequence of fair bets and M_n is the amount of money s/he has after n bets then M_n is a martingale – even if the bets made depend on the outcomes of previous bets, that is, even if the gambler plays a strategy.

Some evidence for some of the above:

Random walk: $U_1, U_2, \ldots$ iid with

$$P(U_i = 1) = P(U_i = -1) = 1/2$$

and $Y_k = U_1 + \cdots + U_k$ with $Y_0 = 0$. Then

$$\begin{aligned}
&\mathrm{E}(Y_n|Y_0, \ldots, Y_k) \\
&\qquad = \mathrm{E}(Y_n - Y_k + Y_k|Y_0, \ldots, Y_k) \\
&\qquad = \mathrm{E}(Y_n - Y_k|Y_0, \ldots, Y_k) + Y_k \\
&\qquad = \sum_{k+1}^{n} \mathrm{E}(U_j|U_1, \ldots, U_k) + Y_k \\
&\qquad = \sum_{k+1}^{n} \mathrm{E}(U_j) + Y_k \\
&\qquad = Y_k
\end{aligned}$$

Things to notice:

Y_k treated as constant given $Y_1, \ldots, Y_k$.

Knowing $Y_1, \ldots, Y_k$ is equivalent to knowing $U_1, \ldots, U_k$.

For $j > k$ we have U_j independent of $U_1, \ldots, U_k$ so conditional expectation is unconditional expectation.

Since Standard Brownian Motion is limit of such random walks we get martingale property for standard Brownian motion.

Poisson Process: $X(t) = N(t) - \lambda t$. Fix $t > s$.

$$\begin{aligned}
&\mathrm{E}(X(t)|X(u); 0 \le u \le s) \\
&\quad = \mathrm{E}(X(t) - X(s) + X(s)|\mathcal{H}_s) \\
&\quad = \mathrm{E}(X(t) - X(s)|\mathcal{H}_s) + X(s) \\
&\quad = \mathrm{E}(N(t) - N(s) - \lambda(t-s)|\mathcal{H}_s) + X(s) \\
&\quad = \mathrm{E}(N(t) - N(s)) - \lambda(t-s) + X(s) \\
&\quad = \lambda(t-s) - \lambda(t-s) + X(s) \\
&\quad = X(s)
\end{aligned}$$

Things to notice:

I used independent increments.

$\mathcal{H}_s$ is shorthand for the conditioning event.

Similar to random walk calculation.

Black Scholes

We model the price of a stock as
$$X(t) = x_0 e^{Y(t)}$$
where
$$Y(t) = \sigma B(t) + \mu t$$
is a Brownian motion with drift (B is standard Brownian motion).

If annual interest rates are $e^{\alpha} - 1$ we call α the instantaneous interest rate; if we invest \$1 at time 0 then at time t we would have $e^{\alpha t}$. In this sense an amount of money $x(t)$ to be paid at time t is worth only $e^{-\alpha t} x(t)$ at time 0 (because that much money at time 0 will grow to $x(t)$ by time t).

Present Value: If the stock price at time t is $X(t)$ per share then the present value of 1 share to be delivered at time t is
$$Z(t) = e^{-\alpha t} X(t)$$
With X as above we see
$$Z(t) = x_0 e^{\sigma B(t) + (\mu - \alpha) t}$$

Now we compute

$$\mathrm{E}\left\{Z(t)|Z(u); 0 \le u \le s\right\} = \mathrm{E}\left\{Z(t)|B(u); 0 \le u \le s\right\}$$

for $s < t$. Write
$$Z(t) = x_0 e^{\sigma B(s) + (\mu - \alpha) t} \times e^{\sigma(B(t) - B(s))}$$
Since B has independent increments we find

$$\mathrm{E}\left\{Z(t)|B(u); 0 \le u \le s\right\} = x_0 e^{\sigma B(s) + (\mu - \alpha) t} \times \mathrm{E}\left[e^{\sigma\{B(t) - B(s)\}}\right]$$

Note: $B(t) - B(s)$ is $N(0, t-s)$; the expected value needed is the moment generating function of this variable at σ.

Suppose $U \sim N(0,1)$. The Moment Generating Function of U is

$$M_U(r) = \mathrm{E}(e^{rU}) = e^{r^2/2}$$

Rewrite

$$\sigma\{B(t) - B(s)\} = \sigma\sqrt{t-s}U$$

where $U \sim N(0,1)$ to see

$$\mathrm{E}\left[e^{\sigma\{B(t)-B(s)\}}\right] = e^{\sigma^2(t-s)/2}$$

Finally we get

$$\begin{aligned}
&\mathrm{E}\{Z(t)|Z(u); 0 \le u \le s\} \\
&= x_0 e^{\sigma B(s)+(\mu-\alpha)s} e^{(\mu-\alpha)(t-s)+\sigma^2(t-s)/2} \\
&= Z(s)
\end{aligned}$$

provided

$$\mu + \sigma^2/2 = \alpha\,.$$

If this identity is satisfied then the present value of the stock price is a martingale.

Option Pricing

Suppose you can pay $\$c$ today for the right to pay K for a share of this stock at time t (regardless of the actual price at time t).

If, at time t, $X(t) > K$ you will **exercise** your **option** and buy the share making $X(t) - K$ dollars.

If $X(t) \le K$ you will not exercise your option; it becomes worthless.

The present value of this option is

$$e^{-\alpha t}(X(t) - K)_+ - c$$

where

$$z_+ = \begin{cases} z & z > 0 \\ 0 & z \le 0 \end{cases}$$

(Called **positive part** of z.)

In a fair market:

- The discounted share price $e^{-\alpha t}X(t)$ is a martingale.
- The expected present value of the option is 0.

So:

$$c = \mathrm{E}\left[e^{-\alpha t}\{X(t) - K\}_+\right]$$

Since

$$X(t) = x_0 e^{N(\mu t, \sigma^2 t)}$$

we are to compute

$$\mathrm{E}\left\{\left(x_0 e^{\sigma t^{1/2}U + \mu t} - K\right)_+\right\}$$

This is

$$\int_a^\infty \left(x_0 e^{bu+d} - K\right) e^{-u^2/2} du/\sqrt{2\pi}$$

where

$$\begin{aligned}
a &= (\log(K/x_0) - \mu t)/(\sigma t^{1/2}) \\
b &= \sigma t^{1/2} \\
d &= \mu t
\end{aligned}$$

Evidently

$$K\int_a^\infty e^{-u^2/2}du/\sqrt{2\pi} = KP(N(0,1) > a)$$

The other integral needed is

$$\begin{aligned}\int_a^\infty e^{-u^2/2+bu}du/\sqrt{2\pi} \\ &= \int_a^\infty \frac{e^{-(u-b)^2/2}e^{b^2/2}}{\sqrt{2\pi}}du \\ &= \int_{a-b}^\infty \frac{e^{-v^2/2}e^{b^2/2}}{\sqrt{2\pi}}dv \\ &= e^{b^2/2}P(N(0,1) > a-b)\end{aligned}$$

Introduce the notation

$$\Phi(v) = P(N(0,1) \le v) = P(N(0,1) > -v)$$

and do all the algebra to get

$$\begin{aligned}c &= \left\{x_0 e^{-\alpha t} e^{b^2/2+d}\Phi(b-a) - Ke^{-\alpha t}\Phi(-a)\right\} \\ &= \left\{x_0 e^{(\mu+\sigma^2/2-\alpha)t}\Phi(b-a) - Ke^{-\alpha t}\Phi(-a)\right\} \\ &= \left\{x_0\Phi(b-a) - Ke^{-\alpha t}\Phi(-a)\right\}\end{aligned}$$

This is the Black-Scholes option pricing formula.

Simulation

Monte Carlo computation of expected value:

To compute $\mathrm{E}(X)$: do experiment n times to generate n independent observations all with same distribution as X.

Get $X_1, \ldots, X_n$.

Use

$$\bar{X} = \frac{1}{n}\sum_1^n X_i$$

as an estimate of $\mathrm{E}(X)$

To estimate $\mathrm{E}(g(X))$: use same X_i and compute

$$\frac{1}{n}\sum_1^n g(X_i)\,.$$

Random Number Generation

In practice: random quantities are not used.

Use: **pseudo-random** uniform random numbers.

They "behave like" iid Uniform(0,1) variables.

Many, many generators. One standard kind: **linear congruential generators**.

Start with x_0 an integer in range $0 \le x_0 < m$. Compute

$$x_{n+1} = ax_n + b \mod m$$

Here the **mod** means compute the remainder after division by m.

Integers a and b are chosen so that

- the sequence $x_0, \ldots, x_{m-1}$ runs through the set $\{0, \ldots, m-1\}$ (or as much of it is possible, sometimes).
- the sequence of pairs
$$(x_0, x_1), (x_1, x_2), \ldots, (x_{m-1}, x_m)$$
fills up as much of the square $\{(i,j) : 1 \leq i, j \leq m-1\}$.
- similarly for triples (x_k, x_{k+1}, x_{k+2}).

Use
$$U_n = \frac{x_n}{m}$$
as Uniform(0,1) random numbers.

General Random Variables

We now pretend we can generate $U_1, U_2, \ldots$ which are iid Uniform(0,1).
Monte Carlo methods for generating a sample:

- Transformation:
  - Inverse Probability Integral Transform
  - Multivariate (e.g. Box Müller)
- Acceptance Rejection
- Markov Chain Monte Carlo

Inverse (Probability Integral) Transform

Fact: F continuous CDF and X, U satisfy
$$U = F(X)$$
then $U \sim$Uniform(0,1) if and only if $X \sim F$.
Proof: For simplicity: assume F strictly increasing on inverval (a, b) with $F(b) = 1$ and $F(a) = 0$.
If $U \sim$Uniform(0,1) then
$$\begin{aligned} P(X \leq x) &= P(F(X) \leq F(x)) \\ &= P(U \leq F(x)) \\ &= F(x) \end{aligned}$$
Conversely: if $X \sim F$ and $0 < u < 1$ then there is a unique x such that $F(x) = u$
$$\begin{aligned} P(U \leq u) &= P(F(X) \leq F(x)) \\ &= P(X \leq x) \\ &= F(x) \\ &= u \end{aligned}$$
Application: generate U and solve $F(X) = U$ for X to get $X = F^{-1}(U)$ which has cdf F.
Example:For the exponential distribution
$$F(x) = 1 - e^{-\lambda x}$$
Set $F(X) = U$ and solve to get
$$X = -\log(1 - U)/\lambda$$
Observation: $U \sim 1 - U$ so $X = -\log(U)/\lambda$ also has an Exponential(λ) distribution.

Example:: for F the standard normal cdf solving

$$F(X) = U$$

requires numerical approximation but such solutions are built in to many languages.

Special purpose transformations:

Example:: If X and Y are iid $N(0,1)$ define $R \ge 0$ and $\Theta \in [0, 2\pi)$ by

$$X = R\cos(\Theta) \qquad Y = R\sin(\Theta)$$

NOTE: Book says

$$\Theta = \tan^{-1}(Y/X)$$

but this takes values in $(-\pi/2, \pi/2)$; notation

$$\tan^{-1}(x, y)$$

means angle θ in $[0, 2\pi)$ so that

$$\frac{x}{\sqrt{x^2+y^2}} = \cos(\theta) \qquad \frac{y}{\sqrt{x^2+y^2}} = \sin(\theta)\,.$$

Then

$$\{(r,\theta) : r \le r^*, \theta \le \theta^*\}$$

is part of circle of radius r^*. (Start at $(r^*, 0)$ and go clockwise to angle θ^*.)

Compute joint cdf of R, Θ at r^*, θ^*.

Define

$$C = \{(x,y) : x^2 + y^2 \le (r^*)^2, \qquad 0 \le \tan^{-1}(x,y) \le \theta^*\}$$

Then

$$\begin{aligned} P(R \le r, &\Theta \le \theta) \\ &= \iint_C \frac{e^{-(x^2+y^2)/2}}{2\pi} dx dy \end{aligned}$$

Do integral in polar co-ordinates to get

$$\int_0^{r^*} \int_0^{\theta^*} \frac{e^{-r^2/2}}{2\pi} r dr d\theta$$

The θ integral just gives

$$\theta^*/(2\pi)$$

The r integral then gives

$$1 - e^{-(r^*)^2/2}$$

So

$$P(R \le r^*, \Theta \le \theta^*) = \left(1 - e^{-(r^*)^2/2}\right) \times \frac{\theta^*}{2\pi}$$

Product of two cdfs so: R and Θ are independent.

Moreover R^2 has cdf

$$P(R^2 \le x) = 1 - e^{-x/2}$$

which is Exponential with rate $1/2$.

Θ is Uniform$(0, 2\pi)$.

So: generate U_1, U_2 iid Uniform(0,1).
Define

$$\Theta = 2\pi U_2$$

and

$$R = \sqrt{-2\log(U_1)}$$

Put

$$X = R\cos(\Theta) \qquad Y = R\sin(\Theta)$$

You have generated two independent $N(0,1)$ variables.

Acceptance Rejection

If $F(X) = U$ difficult to solve (eg F hard to compute) sometimes use acceptance rejection.
Goal: generate $X \sim F$ where $F' = f$.
Tactic: find density g and constant c such that

$$f(x) \le cg(x)$$

and s.t. can generate Y from density g.
Algorithm:

1. Generate Y which has density g.
2. Generate U ~Uniform(0,1) independent of Y.
3. If $U \le f(Y)/\{cg(Y)\}$ let $X = Y$.
4. If not reject Y and go back to step 1; repeat, generating new Y and U (independently) till you accept a Y.

Facts:

- Since you repeat the same experiment each trial: number of trials, N, success is Geometric (and sure to be finite).
- The probability of success on a individual trial is

$$\begin{aligned} P[U &\le f(Y)/\{cg(Y)\}] \\ &= \int P[U \le f(y)/\{cg(y)\}|Y = y]g(y)dy \\ &= \int \frac{f(y)}{cg(y)}g(y)dy \\ &= \frac{1}{c}\int f(y)dy \\ &= 1/c \end{aligned}$$

so

$$\mathrm{E}(N) = 1/c$$

Most important fact: $X \sim f$:
Proof: this is like the old craps example:
We compute

$$P(X \le x) = P[Y \le x | U \le f(Y)/\{cg(Y)\}]$$

(condition on the first iteration where the condition is met).

Condition on Y to get

$$\begin{aligned}
&P(X \le x) \\
&= \frac{\int_{-\infty}^{x} P[U \le f(Y)/\{cg(Y)\} | Y = y] g(y) dy}{P[U \le f(Y)/\{cg(Y)\}]} \\
&= \frac{\int_{-\infty}^{x} \frac{f(y)}{cg(y)} g(y) dy}{1/c} \\
&= \int_{-\infty}^{x} f(y) dy \\
&= F(x)
\end{aligned}$$

as desired.

Example:Half normal distribution: X has density

$$f(x) = \frac{2e^{-x^2/2}}{\sqrt{2\pi}} 1(x > 0)$$

(Density of absolute value of $N(0,1)$.)

Use $g(x) = ae^{-ax}1(x > 0$. To find c maximize

$$\frac{f(x)}{g(x)} = \frac{e^{-x^2/2+ax}}{a\sqrt{2\pi}}$$

Rewrite as

$$\frac{e^{a^2/2} e^{-(x-a)^2/2}}{a\sqrt{2\pi}}$$

Maximum is at $x = a$ giving

$$c = \frac{e^{a^2/2}}{a\sqrt{2\pi}}$$

Choose a to minimize c (or $\log(c)$); get $a = 1$ so

$$c = \frac{e}{\sqrt{2\pi}}$$

Algorithm is then: generate $Y = -\log(U_1)$ and then compute

$$e^{-(Y-1)^2}$$

Generate U_2 and if

$$U_2 \le e^{-(Y-1)^2}$$

accept Y otherwise try again.

To generate $N(0,1)$: use a third U to pick a sign at random: negative if $u < 1/2$ otherwise positive.

Discrete Distributions

The inverse transformation method works for discrete distributions.

X has possible values $\{x_1, x_2, \ldots\}$ with probabilities $\{p_1, p_2, \ldots\}$ compute cumulative probabilities

$$P_k = p_1 + \cdots + p_k$$

with $P_0 = 0$.

Generate U ~Uniform(0,1).

Find k such that

$$P_{k-1} < U \le P_k$$

Put $X = x_k$.