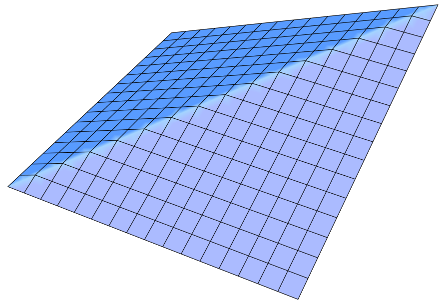

STAT 333: Applied Probability

Prof. Diana Skrzydlo • Spring 2014 • University of Waterloo

MW 4:30–5:20PM in MC 2065 and R 11:30AM–12:20PM in DC 1351

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is https://rob-zimmerman.github.io/coursenotes.

Contents

Lecture 1 — Monday, May 05, 2014

1. Probability Foundations

Experiments, Sample Spaces, and Events

We start by recalling some familiar concepts. An **experiment** is some process with a number of possible outcomes and a **trial** is one iteration of an experiment. An **outcome** is one potential result in an experiment, and a **sample space** is a set of outcomes. An experiment can have multiple sample spaces defined on it. These are not mathematical definitions, but they provide an intuitive grounding for what follows.

From a mathematical perspective, a sample space is simply a set S, and an outcome s is an element of S. We equip S with a σ-algebra $\mathcal{F}$ of measurable subsets.

> **Definition 1.1** An **event** is a measurable subset of the sample space; that is, an element of $\mathcal{F}$.

If $A = \{s_1, s_2, s_3\} \subseteq S$, then we say "$A$ occurs" if the outcome of the experiment is s_1, s_2, or s_3. Otherwise, we say "A does not occur".

Lecture 2 — Wednesday, May 07, 2014

Probability Measures and Bayes' Rule

In this lecture, we formalize probability measures on sample spaces and derive basic identities that we will use repeatedly.

Definition 2.1 We call a sample space S **discrete** if it is countable, including finite. An uncountable sample space is often called **continuous**.

Events belong to $\mathcal{F}$. We say two events A and B are **mutually exclusive** when they cannot both occur in a single trial, which means that

$$A \cap B = \emptyset.$$

Events can be combined with set operators such as $A \cup B$ and $A \cap B$.

Definition 2.2 A **probability measure** is a function

$$\mathbb{P} : \mathcal{F} \to [0, 1],$$

satisfying the following:

1. $\mathbb{P}(S) = 1$.
2. $\mathbb{P}(A) \geq 0$ for every $A \in \mathcal{F}$.
3. If $A_1, A_2, \ldots$ are mutually exclusive, then
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$
(**countable additivity**).

Proposition 2.3 The following hold:

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ (**inclusion–exclusion**).
3. If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$ (**monotonicity**).
4. For any events $A_1, A_2, \ldots,$
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

(**countable subadditivity**).

Proof. First, $\mathbb{P}(\emptyset) = 0$: apply countable additivity to the disjoint sequence $S, \emptyset, \emptyset, \ldots$, whose union is S. Finite additivity for disjoint events follows by padding a finite collection with empty sets.

For (1), since $A \cup A^c = S$ and $A \cap A^c = \emptyset$,

$$\mathbb{P}(S) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1.$$

Therefore,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

For (2), Write $A \cup B = A \cup (B \setminus A)$, where the union is disjoint. Hence

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A).$$

Also, $B = (B \setminus A) \cup (A \cap B)$, again disjoint, so

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

Rearranging gives $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Substituting yields

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

For (3),

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

For (4), Let

$$B_1 = A_1, \qquad B_2 = A_2 \setminus A_1, \qquad B_3 = A_3 \setminus (A_1 \cup A_2), \; \ldots$$

Then the B_i are mutually exclusive, and

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i.$$

By countable additivity,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \sum_{i=1}^{\infty} \mathbb{P}\left(A_i \setminus (A_1 \cup \cdots \cup A_{i-1})\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

$\square$

Definition 2.4 Two events A and B are called **independent** when

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

More generally, events $A_1, \ldots, A_k$ are independent when every finite subcollection factorizes: for each nonempty $J \subseteq \{1, \ldots, k\}$,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Definition 2.5 The **conditional probability** of A given B is

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

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

If A and B are independent, then

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

Rearranging the definition of conditional probability gives

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

which yields Bayes' rule:

Theorem 2.6 (Bayes' rule) Let A and B be events with $0 < \mathbb{P}(A) < 1$ and $\mathbb{P}(B) > 0$.

Then

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B \mid A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B \mid A)}{\mathbb{P}(A)\mathbb{P}(B \mid A) + \mathbb{P}(A^c)\mathbb{P}(B \mid A^c)}.$$

Example 2.7 Consider three doors, one hiding a car and two hiding goats. The contestant chooses door 1. Monty then opens one of the other two doors, always revealing a goat, and if Monty has a choice he picks uniformly at random.

Suppose Monty opens door 3. Let A be the event that the car is behind door 2, and let B be the event that Monty opens door 3. Then

$$\mathbb{P}(A) = \frac{1}{3} \qquad \text{and} \qquad \mathbb{P}(B \mid A) = 1,$$

since if the car is behind door 2 then Monty must open door 3.

To find $\mathbb{P}(B)$, split according to the location of the car:

$$\begin{aligned}\mathbb{P}(B) = \mathbb{P}(B \mid \text{car behind door } 1)&\mathbb{P}(\text{car behind door } 1)\\ + \mathbb{P}(B \mid \text{car behind door } 2)&\mathbb{P}(\text{car behind door } 2)\\ + \mathbb{P}(B \mid \text{car behind door } 3)&\mathbb{P}(\text{car behind door } 3).\end{aligned}$$

Hence

$$\mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}.$$

Therefore, by Bayes' rule,

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

So once Monty opens door 3, the probability that the car is behind door 2 is $\frac{2}{3}$, so the contestant should switch.

Lecture 3 — Thursday, May 08, 2014

2. Random Variables and Distributions

Random Variables and Discrete Distributions

In this lecture, we shift from probabilities of events to random variables and distributions.

Definition 3.1 A **random variable** X is a measurable function from the sample space to $\mathbb{R}$.

Measurability depends on the chosen σ-algebra on S; it ensures that events such as $\{X \le x\}$ have probabilities.

Example 3.2 The experiment is to roll 2 dice. Then

$$S = \{(i, j) \mid 1 \le i, j \le 6\}.$$

A possible random variable is

$$X(i, j) = i + j,$$

Then

$$X : S \to \{2, \ldots, 12\}.$$

Definition 3.3 A random variable X is **discrete** if some countable set C satisfies $\mathbb{P}(X \in C) = 1$. It is **continuous** (more precisely, absolutely continuous) if its distribution has a density with respect to Lebesgue measure.

Discrete Random Variables

We now define the main functions associated with a discrete random variable and record standard model examples.

Definition 3.4 The **probability mass function (pmf)** for a random variable X is the function $k \mapsto \mathbb{P}(X = k)$, for all k in the range of X. We must have

$$\mathbb{P}(X = k) \geq 0 \qquad \text{and} \qquad \sum_k \mathbb{P}(X = k) = 1,$$

where each event "$X = k$" is mutually exclusive and they cover the entire sample space.

Definition 3.5 The **cumulative distribution function (cdf)** for a random variable X is

$$F_X(k) = \mathbb{P}(X \leq k) = \sum_{j \leq k} \mathbb{P}(X = j).$$

Definition 3.6 A **Bernoulli trial** is a trial such that the outcome is either a success (S) or a failure (F).

We have the following resulting random variables.

Definition 3.7 Let

$$\mathbb{1}_j = \begin{cases} 1 & \text{if trial } j \text{ is S,} \\ 0 & \text{if trial } j \text{ is F.} \end{cases}$$

For $p \in [0, 1]$, $\mathbb{1}_j$ is a **Bernoulli (indicator)** random variable with

$$\mathbb{P}(\mathbb{1}_j = 1) = p \qquad \text{and} \qquad \mathbb{P}(\mathbb{1}_j = 0) = 1 - p.$$

Definition 3.8 Let $n \in \mathbb{Z}_{\geq 0}$ and $p \in [0, 1]$. If X is the number of successes in n independent Bernoulli trials with common success probability p, then X has a **binomial distribution** with

$$\text{Range}(X) = \{0, 1, 2, \ldots, n\} \qquad \text{and} \qquad \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

If $n = 1$, then X is just a Bernoulli random variable. We can express $X \sim \text{Bin}(n, p)$ as the sum of n Bernoulli(p) random variables:

$$X = \sum_{j=1}^{n} \mathbb{1}_j.$$

Definition 3.9 Let $p \in [0,1]$, and let Y be the number of independent Bernoulli trials before the first success, inclusive, with $Y = \infty$ if no success ever occurs. Then Y takes values in $\{1, 2, \dots\} \cup \{\infty\}$, and for finite k,

$$\mathbb{P}(Y = k) = p(1-p)^{k-1}.$$

The statement "$Y = \infty$" means that S never occurs.

Definition 3.10 A nonnegative extended random variable is **proper** when $\mathbb{P}(Y = \infty) = 0$, or equivalently, when $\mathbb{P}(Y < \infty) = 1$.

Proposition 3.11 A Geo(p) random variable is proper if $p > 0$.

Proof.

$$\mathbb{P}(Y = \infty) = \lim_{n \to \infty} \mathbb{P}(Y > n) = \lim_{n \to \infty} (1-p)^n = \begin{cases} 1, & p = 0, \\ 0, & p > 0. \end{cases}$$

So it is proper when p is positive. $\square$

Definition 3.12 Let $r \in \mathbb{Z}^+$ and $p \in (0,1]$. If Z is the number of independent Bernoulli trials required to obtain r successes, including the final success, then Z has a **negative binomial distribution** with

$$\text{Range}(Z) = \{r, r+1, r+2, \dots\} \qquad \text{and} \qquad \mathbb{P}(Z = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

If $r = 1$, then $Z \sim \text{Geo}(p)$.

The geometric and negative binomial distributions are both **waiting-time** distributions. Moreover, if $Z \sim \text{NegBin}(r, p)$, then

$$Z = \sum_{j=1}^{r} Y_j \qquad \text{and} \qquad Y_j \stackrel{\text{iid}}{\sim} \text{Geo}(p).$$

Lecture 4 — Monday, May 12, 2014

We now transition from discrete models to the continuous distributions that arise from Poisson process waiting times.

Informally, a **Poisson process** counts events that occur randomly and uniformly through time or space, at rate λ. We will give several rigorous definitions later in the course.

Definition 4.1 Let X be the number of events in a time interval of length $t \geq 0$ in a Poisson process of rate $\lambda > 0$. Then X has a **Poisson distribution**. We write

$$X \sim \text{Poisson}(\lambda t).$$

Then

$$\text{Range}(X) = \{0, 1, 2, \dots\} \qquad \text{and} \qquad \mathbb{P}(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

This model is useful for rare events.

Continuous Random Variables

For continuous random variables, probabilities are computed from areas under densities rather than from point masses. With a continuous random variable X, we have

$$\mathbb{P}(X = k) = 0$$

for every $k \in \mathbb{R}$, because a singleton has Lebesgue measure zero and probabilities are integrals of the density.

Definition 4.2 Let X be a continuous random variable. The **probability density function (pdf)** of X is a function $f(x)$ which satisfies

$$\mathbb{P}(a < X < b) = \int_a^b f(x)\,\mathrm{d}x.$$

Note that $f(x)$ is not itself a probability. Rather,

$$f(x)\,\varepsilon \approx \mathbb{P}\left(x - \frac{\varepsilon}{2} < X < x + \frac{\varepsilon}{2}\right),$$

the probability of X being within ε of x. If f is continuous at x, the ratio of the probability on the right to ε converges to $f(x)$ as $\varepsilon \to 0$.

Proposition 4.3 If $f(x)$ is the pdf of a continuous random variable X, then the following hold:

1. $$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

2. $$f(x) \geq 0 \quad \text{for all } x.$$

3. The cdf satisfies
$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$
Moreover, $F'(x) = f(x)$ for almost every x.

We also have the following continuous random variables.

Definition 4.4 If $X \sim \text{Unif}(a, b)$, where $a < b$, then X has a **uniform distribution** with

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b, \\ 0, & \text{otherwise}, \end{cases} \qquad \text{and} \qquad F_X(x) = \begin{cases} 0, & x \leq a, \\ \dfrac{x-a}{b-a}, & a < x < b, \\ 1, & x \geq b. \end{cases}$$

Most commonly, $X \sim \text{Unif}(0, 1)$, so $f_X(x) = 1$ and $F_X(x) = x$ for $0 < x < 1$.

Definition 4.5 If $X \sim \text{Exp}(\lambda)$, where $\lambda > 0$, then X has an **exponential distribution** with

$$\text{Range}(X) = [0, \infty), \qquad f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$
$$\text{and} \qquad F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

For $x \geq 0$, the **survival function** of an Exp(λ) random variable is

$$\mathbb{P}(X > x) = 1 - F(x) = e^{-\lambda x}.$$

Here X represents the waiting time until the next event in a Poisson process. The following result is interesting:

Proposition 4.6 (Memoryless property) If $X \sim \text{Exp}(\lambda)$, then for all $s, t \geq 0$,

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t).$$

Proof.

$$\mathbb{P}(X > s + t \mid X > s) = \frac{\mathbb{P}(X > s + t, X > s)}{\mathbb{P}(X > s)} = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \frac{e^{-(s+t)\lambda}}{e^{-s\lambda}} = e^{-\lambda t} = \mathbb{P}(X > t).$$

□

Thus, conditional on having waited s time units already, the remaining wait has the same distribution as the original one.

Definition 4.7 If $X \sim \text{Gamma}(\alpha, \lambda)$, where $\alpha > 0$ and $\lambda > 0$, using the shape–rate convention, then X has a **gamma distribution** with

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x > 0, \qquad \text{and} \qquad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, \mathrm{d}x.$$

When $\alpha \in \mathbb{Z}^+$, we have $\Gamma(\alpha) = (\alpha - 1)!$. In that case, we can interpret X as the total waiting time in a Poisson process until α events occur.

Definition 4.8 If $X \sim \mathcal{N}(\mu, \sigma^2)$, where $\sigma > 0$, then X has a **normal distribution** with

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}.$$

In this course we have no interest in the normal distribution, so it is okay to forget about it forever.

This table summarizes the analogous quantities for Bernoulli trials and Poisson processes.

process	Bernoulli trials	Poisson processes at rate λ
single trial	Bernoulli(p)	
counting random variable	$\text{Bin}(n,p)$ (n trials)	Poisson(λt) (t time)
waiting time (single event)	Geo(p)	Exp(λ)
waiting for multiple events	NegBin(r,p)	Gamma(α,λ)

Lecture 5 — Wednesday, May 14, 2014

3. Expectation and Joint Distributions

Expectation, Variance, and Joint Distributions

We now summarize expectation, variance, and basic multivariate constructions that will be used throughout the course.

Intuitively, the expectation of a random variable X is the mean (average) of X.

Definition 5.1 For a discrete random variable X with $\mathbb{E}[|X|] < \infty$, its **expectation** is

$$\mu = \mathbb{E}[X] \coloneqq \sum_{\text{all } x} x \cdot \mathbb{P}(X = x)$$

For an absolutely continuous random variable with $\int_{-\infty}^{\infty} |x| f_X(x)\,\mathrm{d}x < \infty$, it is

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x.$$

in the continuous case.

For a function $g : \mathbb{R} \to \mathbb{R}$ for which the displayed expectation exists, the law of the unconscious statistician gives

$$\mathbb{E}[g(X)] = \sum_x g(x)\mathbb{P}(X = x)$$

in the discrete case, and

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$$

in the continuous case. If g is linear (that is, $g(x) = ax + b$), then

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

Definition 5.2 If $\mathbb{E}[X^2] < \infty$, the **variance** of X is

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

We have the standard identity

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

If $g(x) = ax + b$, then

$$\mathrm{Var}(g(X)) = \mathrm{Var}(ax + b) = a^2 \cdot \mathrm{Var}(X).$$

Definition 5.3 The **joint pmf** for two discrete random variables X, Y is

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

For continuous X, Y, the **joint pdf** of X, Y is

$$f_{X,Y}(x, y).$$

Definition 5.4 The **marginal pmf** of X is

$$\mathbb{P}(X = x) = \sum_{\text{all } y} \mathbb{P}(X = x, Y = y).$$

In the continuous case, the marginal pdf of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y.$$

If X and Y are both discrete and $g : \mathbb{R}^2 \to \mathbb{R}$ is a function, then

$$\mathbb{E}[g(X, Y)] = \sum_{\text{all } x} \sum_{\text{all } y} g(x, y) \cdot \mathbb{P}(X = x, Y = y)$$

If X and Y are both continuous, then

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

These generalize in the natural way.

Definition 5.5 If X and Y have finite second moments, the **covariance** of X, Y is

$$\mathrm{Cov}(X,Y) := \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\cdot\mathbb{E}[Y].$$

Definition 5.6 Two random variables X, Y are **independent** when, for all x, y,

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

in the discrete case, and

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

in the continuous case.

We have the following important consequence of independence:

Proposition 5.7 If X, Y are independent and the displayed expectations exist, then

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

In particular,

$$\mathrm{Cov}(X,Y) = 0.$$

However, the converse does not necessarily hold.

Proposition 5.8 Suppose

$$Y = a_1X_1 + \cdots + a_nX_n,$$

where the X_i have finite second moments, are not necessarily independent, and the a_i are all in $\mathbb{R}$. Then

$$\mathbb{E}[Y] = \sum_{i=1}^{n} a_i \cdot \mathbb{E}[X_i].$$

Also,

$$\mathrm{Var}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \mathrm{Var}(X_i) + \sum_i \sum_{j \neq i} a_i a_j \cdot \mathrm{Cov}(X_i, X_j).$$

Example 5.9 Roll a six-sided die n times. Let X be the number of faces that have not yet appeared. Find $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Solution. Rather than trying to find the pmf directly, define

$$\mathbb{1}_i = \begin{cases} 1 & \text{if face } i \text{ has not been rolled,} \\ 0 & \text{otherwise,} \end{cases}, \quad i = 1, \ldots, 6.$$

Then

$$X = \sum_{i=1}^{6} \mathbb{1}_i.$$

Indicator random variables have convenient properties:

$$\begin{aligned} \mathbb{E}[\mathbb{1}_i] &= 0 \cdot \mathbb{P}(\mathbb{1}_i = 0) + 1 \cdot \mathbb{P}(\mathbb{1}_i = 1) \\ &= \mathbb{P}(\mathbb{1}_i = 1) \\ &= \mathbb{P}(\text{face } i \text{ has not been rolled in } n \text{ rolls}) =: p. \end{aligned}$$

Also,

$$\mathbb{E}\left[\mathbb{1}_i^2\right] = \mathbb{P}(\mathbb{1}_i = 1) \qquad \text{and} \qquad \mathrm{Var}(\mathbb{1}_i) = \mathbb{P}(\mathbb{1}_i = 1) - \mathbb{P}(\mathbb{1}_i = 1)^2 = p(1-p).$$

Now,

$$\mathbb{P}(\text{face } i \text{ is not rolled in } n \text{ rolls}) = \left(\frac{5}{6}\right)^n.$$

Then

$$\mathbb{E}[X] = \sum \mathbb{E}[\mathbb{1}_i] = \sum_{i=1}^{6} \left(\frac{5}{6}\right)^n = 6 \cdot \frac{5^n}{6^n} = \frac{5^n}{6^{n-1}}.$$

For the variance, first note that for $i \neq j$,

$$\mathbb{E}[\mathbb{1}_i \mathbb{1}_j] = \mathbb{P}(\mathbb{1}_i = 1, \mathbb{1}_j = 1) = \left(\frac{4}{6}\right)^n,$$

so

$$\mathrm{Cov}(\mathbb{1}_i, \mathbb{1}_j) = \left(\frac{4}{6}\right)^n - \left(\frac{5}{6}\right)^{2n}.$$

Therefore,

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{i=1}^{6} \mathrm{Var}(\mathbb{1}_i) + \sum_{i} \sum_{j \neq i} \mathrm{Cov}(\mathbb{1}_i, \mathbb{1}_j) \\ &= 6\left(\left(\frac{5}{6}\right)^n - \left(\frac{5}{6}\right)^{2n}\right) + 30\left(\left(\frac{4}{6}\right)^n - \left(\frac{5}{6}\right)^{2n}\right).\end{aligned}$$

□

Lecture 6 — Thursday, May 15, 2014

Properties of Random Variables

Recall that a random variable Y is proper when $\mathbb{P}(Y = \infty) = 0$, or equivalently, when $\mathbb{P}(Y < \infty) = 1$. In Proposition 3.11, we proved that if $Y \sim \mathrm{Geo}(p)$ and $p > 0$, then Y is proper. We could also prove it by showing that $\mathbb{P}(Y < \infty) = 1$:

Alternative proof of Proposition 3.11.

$$\mathbb{P}(Y < \infty) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(Y = k) = \lim_{n\to\infty} \sum_{k=1}^{n} p(1-p)^{k-1} = \lim_{n\to\infty} \frac{p(1-(1-p)^n)}{1-(1-p)} = 1.$$

Hence $\mathbb{P}(Y = \infty) = 0$, so Y is proper. □

There is yet another way to prove Proposition 3.11, by first computing

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{k=1}^{\infty} k \cdot \mathbb{P}(Y = k) \\ &= p + 2p(1-p) + 3p(1-p)^2 + 4p(1-p)^3 + \dots \\ &= \frac{p}{1-(1-p)} + \frac{p(1-p)}{1-(1-p)} + \frac{p(1-p)^2}{1-(1-p)} + \dots \\ &= \frac{1}{p} < \infty\end{aligned}$$

and then using the following result:

Proposition 6.1 If Y is a nonnegative extended random variable and $\mathbb{E}[Y] < \infty$, then $\mathbb{P}(Y = \infty) = 0$.

Proof. For every $M > 0$,

$$\mathbb{E}[Y] \geq M\mathbb{P}(Y \geq M) \geq M\mathbb{P}(Y = \infty).$$

If $\mathbb{P}(Y = \infty) > 0$, the right-hand side is unbounded as $M \to \infty$, contradicting $\mathbb{E}[Y] < \infty$. □

However, the converse does not necessarily hold. Nonnegative extended random variables can be classified by their means:

1. Proper + $\mathbb{E}[Y] < \infty$ is "**short proper**".
2. Proper + $\mathbb{E}[Y] = \infty$ is "**null proper**".
3. Improper always has $\mathbb{E}[Y] = \infty$.

Example 6.2 A mouse is in the center of a maze with four doors. If it leaves through the i-th door, it returns in i minutes for $i = 1, 2, 3$. Through door 4, it escapes after 4 minutes. Each time it returns, it chooses a new door randomly with no memory.

Let Z be the number of attempts needed to escape. Then $Z \sim \text{Geo}(1/4)$, so Z is short proper. Let X be the total time to escape the maze; this is also short proper. Finally, let Y

be the number of minutes to return to the center. Since

$$\mathbb{P}(Y = \infty) = 1/4 \neq 0,$$

Y is improper.

Example 6.3 Let X be defined on $\mathbb{N}$ by

$$\mathbb{P}(X = k) = \frac{c}{k^2}.$$

Choose c so that X is proper, and classify its propriety.

Solution. We need

$$c \sum_{k=1}^{\infty} \frac{1}{k^2} = 1 \implies c = \frac{6}{\pi^2}.$$

Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot \frac{6/\pi^2}{k^2} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

Hence X is null proper. □

Lecture 7 — Wednesday, May 21, 2014

4. Conditioning

Conditional Random Variables and Expectation

If X and Y are discrete proper random variables, then for each fixed value y with $\mathbb{P}(Y = y) > 0$, we can define a new random variable $X \mid Y = y$.

Definition 7.1 For $x \in \text{Range}(X \mid Y = y)$, the **conditional pmf** of $X \mid Y = y$ is

$$\mathbb{P}(X = x \mid Y = y) := \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

We can think of this as fixing $Y = y$ and allowing X to vary over its conditional range.

Proposition 7.2 If $\mathbb{P}(Y = y) \neq 0$, then $\mathbb{P}(X = x \mid Y = y)$ is a valid proper pmf.

Proof. Nonnegativity is immediate:

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \geq 0.$$

Also,

$$\begin{aligned}
\sum_{\text{all } x} \mathbb{P}(X = x \mid Y = y) &= \sum_{\text{all } x} \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \\
&= \frac{1}{\mathbb{P}(Y = y)} \sum_{\text{all } x} \mathbb{P}(X = x, Y = y) \\
&= \frac{\mathbb{P}(Y = y)}{\mathbb{P}(Y = y)} \\
&= 1.
\end{aligned}$$

Hence this is a proper pmf. $\square$

Definition 7.3 If X is discrete, the **conditional expectation** of $X \mid Y = y$ is

$$\mathbb{E}[X \mid Y = y] := \sum_{\text{all } x} x \cdot \mathbb{P}(X = x \mid Y = y).$$

The conditional expectation obeys the same linearity rule as ordinary expectation:

$$\mathbb{E}[aX + b \mid Y = y] = a \cdot \mathbb{E}[X \mid Y = y] + b.$$

Proposition 7.4 If X and Y are independent, then $\mathbb{E}[X \mid Y = y] = \mathbb{E}[X]$.

Proof.

$$\begin{aligned}
\mathbb{E}[X \mid Y = y] &= \sum_{\text{all } x} x \cdot \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \\
&= \sum_{\text{all } x} x \cdot \frac{\mathbb{P}(X = x)\mathbb{P}(Y = y)}{\mathbb{P}(Y = y)} \\
&= \sum_{\text{all } x} x \cdot \mathbb{P}(X = x) \\
&= \mathbb{E}[X].
\end{aligned}$$

□

Note that the converse need not hold.

Example 7.5 Let $t = m + n$ Bernoulli trials with $m, n \geq 0$ and $0 < p < 1$. Let X count the successes in all t trials, and let Y count the successes in the first m trials. Fix $y \in \{0, \ldots, m\}$. Find the conditional pmf and the conditional expectation of $X \mid Y = y$.

Solution. Observe

$$\text{Range}(X \mid Y = y) = \{y, y + 1, \ldots, y + n\}.$$

Let

$$Z := \text{number of successes in the last } n \text{ trials},$$

so Z is independent of Y. For $x \in \text{Range}(X \mid Y = y)$,

$$\begin{aligned}
\mathbb{P}(X = x \mid Y = y) &= \frac{\mathbb{P}(Z = x - y, Y = y)}{\mathbb{P}(Y = y)} \\
&= \frac{\mathbb{P}(Z = x - y)\mathbb{P}(Y = y)}{\mathbb{P}(Y = y)} \\
&= \mathbb{P}(Z = x - y) \\
&= \binom{n}{x - y} p^{x-y}(1 - p)^{n-x+y}.
\end{aligned}$$

Therefore

$$\mathbb{E}[X \mid Y = y] = \sum_{x=y}^{y+n} x \binom{n}{x - y} p^{x-y}(1 - p)^{n-x+y}.$$

Setting $u = x - y$, we get

$$\mathbb{E}\left[X \mid Y = y\right] = \sum_{u=0}^{n}(u+y)\binom{n}{u}p^u(1-p)^{n-u} = np + y,$$

which makes intuitive sense. □

The conditional range and conditional expectation are usually functions of y. If y is not fixed and instead varies according to the pmf of Y, then we obtain a new random variable.

Definition 7.6 Let $g(y) = \mathbb{E}[X \mid Y = y]$. The **conditional expectation** given Y is the random variable

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

Thus it takes the value $g(y)$ whenever $Y = y$; distinct values of y may, of course, give the same value of $g(y)$.

Lecture 8 — Thursday, May 22, 2014

Recall that for $X \mid Y = y$, we defined $\mathbb{P}(X = x \mid Y = y)$, $\mathbb{E}\left[X \mid Y = y\right]$, and the random variable $\mathbb{E}\left[X \mid Y\right]$, which is a function of Y.

Theorem 8.1 (Double averaging (discrete case)) For discrete random variables X and Y, if $X \geq 0$ or $\mathbb{E}[|X|] < \infty$, then

$$\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \mathbb{E}\left[X\right].$$

Proof.

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] &= \sum_y \mathbb{E}\left[X \mid Y = y\right]\mathbb{P}(Y = y)\\ &= \sum_y \sum_x x\mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y)\\ &= \sum_x x \sum_y \mathbb{P}(X = x, Y = y)\end{aligned}$$

$$= \sum_x x\mathbb{P}(X = x) = \mathbb{E}[X].$$

□

Example 8.2 Let X count the successes in t trials and let Y count the successes in the first m trials. We obtained

$$\mathbb{E}[X \mid Y = y] = np + y.$$

Therefore $\mathbb{E}[X \mid Y] = np + Y$, and

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[np + Y] = np + \mathbb{E}[Y] = np + mp = p(n + m) = pt = \mathbb{E}[X].$$

This agrees with the direct calculation from $X \sim \text{Bin}(t, p)$.

These techniques are especially useful when $\mathbb{E}[X]$ is difficult to compute directly or when the distribution of X is unknown.

Example 8.3 Consider the mouse maze in Figure 1.

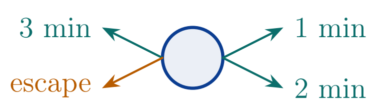

FIGURE 1

Let X be the number of minutes until escape. Find $\mathbb{E}[X]$.

Solution. Condition on the first door D the mouse chooses, where

$$\mathbb{P}(D = d) = \frac{1}{4}, \quad d = 1, \ldots, 4.$$

We have

$$X \mid (D = 4) = 4, \qquad X \mid (D = 1) \overset{\text{d}}{=} 1 + X',$$
$$X \mid (D = 2) \overset{\text{d}}{=} 2 + X' \qquad \text{and} \qquad X \mid (D = 3) \overset{\text{d}}{=} 3 + X',$$

where X' is an independent copy of X. Since the mouse has no memory,

$$\begin{aligned}
\mathbb{E}[X \mid D = 1] &= 1 + \mathbb{E}[X], \\
\mathbb{E}[X \mid D = 2] &= 2 + \mathbb{E}[X], \\
\mathbb{E}[X \mid D = 3] &= 3 + \mathbb{E}[X], \\
\mathbb{E}[X \mid D = 4] &= 4.
\end{aligned}$$

Therefore,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid D]] = \frac{1}{4}(1 + \mathbb{E}[X]) + \frac{1}{4}(2 + \mathbb{E}[X]) + \frac{1}{4}(3 + \mathbb{E}[X]) + \frac{1}{4}(4),$$

so $\mathbb{E}[X] = 10$. This is consistent with the geometric trial interpretation: the expected number of tries is 4, with 4 minutes on the successful try and 2 minutes on average on each unsuccessful try. □

Example 8.4 Consider Bernoulli(p) trials with $0 < p \leq 1$, and let $k \geq 1$. Let A_k be the event that we see k successes in a row. We are interested in the expectation of

$$T_{A_k},$$

the number of trials needed to obtain A_k. For $k = 1$,

$$T_{A_1} \sim \text{Geo}(p) \qquad \text{and} \qquad \mathbb{E}[T_{A_1}] = \frac{1}{p}.$$

For $k > 1$, let $T_{A_k|A_{k-1}}$ be the number of trials needed to see k successes, given that we have just seen $k - 1$ in a row. Then

$$T_{A_k} = T_{A_{k-1}} + T_{A_k|A_{k-1}}.$$

We continue this calculation next lecture.

Lecture 9 — Monday, May 26, 2014

Example 9.1 Continuing Example 8.4, we had

$$T_{A_k} = T_{A_{k-1}} + T_{A_k|A_{k-1}}.$$

To find $\mathbb{E}\left[T_{A_k|A_{k-1}}\right]$, condition on the next trial:

$$\mathbb{E}\left[T_{A_k|A_{k-1}} \mid S\right] = 1 \qquad \text{and} \qquad \mathbb{E}\left[T_{A_k|A_{k-1}} \mid F\right] = \mathbb{E}\left[1 + T_{A_k}\right] = 1 + \mathbb{E}\left[T_{A_k}\right].$$

By double averaging,

$$\mathbb{E}\left[T_{A_k|A_{k-1}}\right] = 1 \cdot p + (1 + \mathbb{E}\left[T_{A_k}\right])(1-p).$$

Hence

$$\mathbb{E}\left[T_{A_k}\right] = \left(\mathbb{E}\left[T_{A_{k-1}}\right] + 1\right)\frac{1}{p}.$$

Since $\mathbb{E}\left[T_{A_1}\right] = \frac{1}{p}$,

$$\mathbb{E}\left[T_{A_2}\right] = \frac{1}{p} + \frac{1}{p^2} \qquad \text{and} \qquad \mathbb{E}\left[T_{A_3}\right] = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3},$$

and in general

$$\mathbb{E}\left[T_{A_k}\right] = \sum_{i=1}^{k} \frac{1}{p^i}.$$

For example, getting black on roulette 10 times in a row takes about 3,339 spins on average.

Conditioning on Continuous Random Variables

Recall that if $f_{X,Y}(x, y)$ is a joint pdf, then the marginal pdfs are

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}x.$$

Definition 9.2 If $X \mid Y = y$ is continuous and $f_Y(y) > 0$, then

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

is the **conditional pdf**, and

$$\mathbb{E}\left[X \mid Y = y\right] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\,\mathrm{d}x$$

is the **conditional expectation**.

Proposition 9.3 (Double averaging (continuous case)) If X and Y are continuous and $\mathbb{E}\left[|X|\right] < \infty$, then

$$\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \int_{-\infty}^{\infty} \mathbb{E}\left[X \mid Y = y\right] f_Y(y)\,\mathrm{d}y = \mathbb{E}\left[X\right].$$

Proof. Because $\mathbb{E}\left[|X|\right] < \infty$, we may interchange integrals by Fubini's theorem. Then

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] &= \int_{-\infty}^{\infty} \mathbb{E}\left[X \mid Y = y\right] f_Y(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\,\mathrm{d}x \right) f_Y(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y \right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x \\
&= \mathbb{E}\left[X\right].
\end{aligned}$$

$\square$

Proposition 9.4 The following hold whenever the displayed conditional expectations exist; for double averaging, it is enough that $X \geq 0$ or $\mathbb{E}[|X|] < \infty$:

1. $$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i \,\middle|\, Y = y\right] = \sum_{i=1}^{n} a_i \cdot \mathbb{E}\left[X_i \mid Y = y\right].$$

2. $$\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \mathbb{E}\left[X\right]$$

   (**double averaging**, or the **law of total expectation**). In particular, if A is an event, then taking $X = \mathbb{1}_A$, we have

   $$\mathbb{P}(A) = \sum_{\text{all } y} \mathbb{P}(A \mid Y = y) \cdot \mathbb{P}(Y = y)$$

   if Y is discrete and

   $$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid Y = y) f_Y(y) \, \mathrm{d}y$$

   if Y is continuous, which is the **law of total probability**.

3. $$\mathbb{E}\left[h(X, Y) \mid Y = y\right] = \mathbb{E}\left[h(X, y) \mid Y = y\right].$$

4. If $X \perp Y$, then
   $$\mathbb{E}\left[g(X) \mid Y = y\right] = \mathbb{E}\left[g(X)\right].$$

Combining (3) and (4), if $X \perp Y$, then

$$\mathbb{E}\left[h(X, Y) \mid Y = y\right] = \mathbb{E}\left[h(X, y) \mid Y = y\right] = \mathbb{E}\left[h(X, y)\right].$$

Lecture 10 — Wednesday, May 28, 2014

Applications of Conditional Expectation

Recall that
$$\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \mathbb{E}\left[X\right],$$
and this identity can be used to compute probabilities and expectations.

If X is discrete and Y is continuous (or vice versa), $\mathbb{E}\left[X \mid Y = y\right]$ may be less direct to interpret.

In practice, either $X \mid Y = y$ has a known distribution whose mean is easy to use, or we rewrite $X \mid Y = y$ in terms of variables independent of Y.

Example 10.1 Suppose $U \sim \text{Unif}(0, 1)$. A coin with $\mathbb{P}(H \mid U = u) = u$ is built, and flipped until the first H appears. Let X be the number of trials until the first H. Find $\mathbb{E}[X]$ and classify its propriety.

Solution. Note that X is not a geometric random variable since U is random; however,

$$X \mid U = u \sim \text{Geo}(u).$$

To find $\mathbb{P}(X > 2)$, condition on $U = u$:

$$\mathbb{P}(X > 2 \mid U = u) = 1 - \mathbb{P}(X = 1 \mid U = u) - \mathbb{P}(X = 2 \mid U = u) = 1 - u - (1-u)u = (1-u)^2.$$

Hence

$$\mathbb{P}(X > 2) = \mathbb{E}\left[\mathbb{P}(X > 2 \mid U)\right] = \mathbb{E}\left[(1-U)^2\right] = \int_0^1 (1-u)^2 f_U(u)\,\mathrm{d}u = \int_0^1 (1-u)^2\,\mathrm{d}u = \frac{1}{3}.$$

Also $\mathbb{E}\left[X \mid U = u\right] = 1/u$, so

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid U\right]\right] = \mathbb{E}\left[\frac{1}{U}\right] = \int_0^1 \frac{1}{u} \cdot f_U(u)\,\mathrm{d}u = \int_0^1 \frac{1}{u}\,\mathrm{d}u = \infty.$$

To classify X, we compute $\mathbb{P}(X = \infty)$. Since $X \mid U = u$ is proper if and only if $u > 0$,

$$\mathbb{P}(X = \infty) = \mathbb{P}(U = 0) = 0.$$

Therefore X is null proper. □

Example 10.2 Three offers are in three sealed envelopes; one is best. We open one at a time and must accept or reject immediately, with no return. Use this strategy: reject the first envelope, accept the second if it is better than the first, and otherwise take the third.

Label the offers A, B, C in increasing order of value, and let A_{win} be the event that the strategy selects the best offer C. The six possible orders are:

A B C   A C B   B A C
B C A   C A B   C B A

Thus

$$\mathbb{P}(A_{\text{win}}) = \frac{1}{2} > \frac{1}{3}.$$

Now generalize to n offers. Reject the first k, record the best among them, then accept the first later offer that exceeds that benchmark. Let A_k be the event that we get the best offer using parameter k. How should we choose k?

This is essentially the secretary problem.

Solution. We want to maximize

$$\mathbb{P}(A_k) = \mathbb{P}(\text{we get the best offer using parameter } k).$$

Condition on X = location of the best offer:

$$\mathbb{P}(A_k) = \sum_{i=1}^{n} \mathbb{P}(A_k \mid X = i)\mathbb{P}(X = i).$$

If $i \le k$, then $\mathbb{P}(A_k \mid X = i) = 0$. If $i > k$, then

$$\mathbb{P}(A_k \mid X = i) = \mathbb{P}(\text{best of } i-1 \text{ is in the first } k) = \frac{k}{i-1}.$$

Therefore

$$\mathbb{P}(A_k) = \sum_{i=k+1}^{n} \frac{k}{i-1} \cdot \frac{1}{n} = \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1}.$$

Write this as

$$\mathbb{P}(A_k) = \frac{k}{n}\big(H_{n-1} - H_{k-1}\big),$$

where $H_n = \sum_{j=1}^{n} 1/j$ is the nth harmonic number. An exact comparison of adjacent choices gives

$$\mathbb{P}(A_{k+1}) - \mathbb{P}(A_k) = \frac{1}{n}\left(\sum_{j=k+1}^{n-1} \frac{1}{j} - 1\right).$$

Thus the success probability increases until the remaining harmonic tail crosses 1, which characterizes an exact maximizing integer k. To obtain its asymptotic scale, use

$$H_n = \ln n + \gamma + o(1),$$

for any sequence $k = k_n$ with $1 \le k_n < n$ and $k_n/n \to x \in (0,1)$, we get

$$\begin{aligned}
\mathbb{P}(A_{k_n}) &= \frac{k_n}{n}\big(H_{n-1} - H_{k_n-1}\big) \\
&= \frac{k_n}{n}\left(\ln\left(\frac{n-1}{k_n-1}\right) + o(1)\right) \\
&= \frac{k_n}{n}\left(\ln\left(\frac{n}{k_n}\right) + o(1)\right).
\end{aligned}$$

Therefore, among sequences $k_n/n \to x \in (0,1)$,

$$\mathbb{P}(A_{k_n}) = -x\ln x + o(1).$$

Define

$$g(x) = x\ln\left(\frac{1}{x}\right) = -x\ln x.$$

Then

$$g'(x) = -\ln x - 1 \qquad \text{and} \qquad g''(x) = -\frac{1}{x} < 0.$$

Setting $g'(x) = 0$ gives $\ln x = -1$, so $x = e^{-1}$, and since $g''(x) < 0$, this is the unique maximizer of the limiting objective. Thus the natural asymptotically optimal choice satisfies

$$\frac{k_n}{n} \to \frac{1}{e}.$$

Taking $k_n = \lfloor n/e \rfloor$, we have $k_n/n = 1/e + o(1)$ and $\ln(n/k_n) = 1 + o(1)$, so

$$\mathbb{P}(A_{k_n}) = \frac{k_n}{n}\left(\ln\left(\frac{n}{k_n}\right) + o(1)\right) = \left(\frac{1}{e} + o(1)\right)(1 + o(1)) = \frac{1}{e} + o(1).$$

$\square$

Lecture 11 — Thursday, May 29, 2014

Computing Variance by Conditioning

Recall that

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

Two common approaches are:

1. Develop an equation for $\mathbb{E}\left[X^2\right]$, often recursively, and
2. Use the conditional variance formula once $X \mid Y = y$ is understood.

Example 11.1 Suppose $X \sim \mathrm{Geo}(p)$, where $0 < p \leq 1$. Find $\mathrm{Var}(X)$ by conditioning on the indicator

$$Y = \mathbb{1}_{\{\text{success on the first trial}\}}.$$

Solution. Since $X \mid Y = 1 = 1$, we have

$$\mathbb{E}\left[X^2 \mid Y = 1\right] = 1.$$

Also, after a first-trial failure, the remaining Bernoulli trials are independent of the first one and have the original law. Thus $X \mid Y = 0$ has the same distribution as $1 + X$, so

$$\begin{aligned}\mathbb{E}\left[X^2 \mid Y = 0\right] &= \mathbb{E}\left[(1+X)^2\right] \\ &= \mathbb{E}\left[1 + 2X + X^2\right] \\ &= 1 + 2\mathbb{E}[X] + \mathbb{E}\left[X^2\right].\end{aligned}$$

By double averaging,

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[\mathbb{E}\left[X^2 \mid Y\right]\right] = 1 \cdot p + (1 + 2\mathbb{E}[X] + \mathbb{E}\left[X^2\right])(1-p) = 1 \cdot p + \left(1 + 2\frac{1}{p} + \mathbb{E}\left[X^2\right]\right)(1-p),$$

so

$$\mathbb{E}\left[X^2\right] = \frac{2-p}{p^2}.$$

Hence

$$\mathrm{Var}(X) = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

□

Definition 11.2 $X \mid Y = y$ has **conditional variance**

$$\mathrm{Var}(X \mid Y = y) = \mathbb{E}\left[(X - \mathbb{E}[X \mid Y = y])^2 \mid Y = y\right] = \mathbb{E}\left[X^2 \mid Y = y\right] - \mathbb{E}[X \mid Y = y]^2$$

and $\mathrm{Var}(X \mid Y)$ is the random variable that takes value $\mathrm{Var}(X \mid Y = y)$ when $Y = y$.

In particular,

$$\mathrm{Var}(X \mid Y) = \mathbb{E}\left[X^2 \mid Y\right] - \mathbb{E}[X \mid Y]^2.$$

Proposition 11.3 If $\mathbb{E}[X^2] < \infty$, then

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y]).$$

Proof. We have

$$\begin{aligned}
\mathbb{E}[\mathrm{Var}(X \mid Y)] &= \mathbb{E}\left[\mathbb{E}\left[X^2 \mid Y\right] - \mathbb{E}[X \mid Y]^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[X^2 \mid Y\right]\right] - \mathbb{E}\left[\mathbb{E}[X \mid Y]^2\right] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[\mathbb{E}[X \mid Y]^2\right].
\end{aligned}$$

Also,

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}[X \mid Y]) &= \mathbb{E}\left[\mathbb{E}[X \mid Y]^2\right] - \mathbb{E}[\mathbb{E}[X \mid Y]]^2 \\
&= \mathbb{E}\left[\mathbb{E}[X \mid Y]^2\right] - \mathbb{E}[X]^2.
\end{aligned}$$

Adding gives

$$\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \mathrm{Var}(X).$$

□

Example 11.4 Toss a coin with $\mathbb{P}(H) = p$ until r heads appear. Let N be the number of tosses required. Then toss a fair coin N times, and let X be the number of heads in those N tosses of the fair coin. Find $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

Solution. We know $N \sim \text{NegBin}(r,p)$, so $\text{Range}(N) = \{r, r+1, \dots\}$. Although X is not binomial, conditionally we have

$$X \mid N = n \sim \text{Bin}\left(n, \frac{1}{2}\right) \qquad \text{and} \qquad \text{Range}(X \mid N = n) = \{0, 1, \dots, n\}.$$

In particular,

$$\mathbb{P}(X = x \mid N = n) = \binom{n}{x}\left(\frac{1}{2}\right)^n,$$

so a direct formula for $\mathbb{P}(X = x)$ is messy. Instead, use conditioning identities.

Since $X \mid N = n \sim \text{Bin}(n, \frac{1}{2})$,

$$\mathbb{E}\left[X \mid N = n\right] = \frac{n}{2} \qquad \text{and} \qquad \text{Var}(X \mid N = n) = \frac{n}{4}.$$

Hence

$$\mathbb{E}\left[X \mid N\right] = \frac{N}{2} \qquad \text{and} \qquad \text{Var}(X \mid N) = \frac{N}{4}.$$

Therefore,

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid N\right]\right] = \mathbb{E}\left[\frac{N}{2}\right] = \frac{1}{2}\mathbb{E}\left[N\right] = \frac{r}{2p}.$$

Also,

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}\left[\text{Var}(X \mid N)\right] + \text{Var}(\mathbb{E}\left[X \mid N\right]) \\
&= \mathbb{E}\left[\frac{N}{4}\right] + \text{Var}\left(\frac{N}{2}\right) \\
&= \frac{1}{4}\mathbb{E}\left[N\right] + \frac{1}{4}\,\text{Var}(N) \\
&= \frac{r}{4p} + \frac{r(1-p)}{4p^2} \\
&= \frac{r}{4p^2}.
\end{aligned}$$

□

Lecture 12 — Monday, June 02, 2014

Compound Random Variables

Definition 12.1 If $X_1, X_2, \ldots$ are independent and identically distributed random variables and N is a nonnegative integer-valued random variable independent of the X_i, then

$$S = \sum_{i=1}^{N} X_i,$$

with $S = 0$ when $N = 0$, is a **compound random variable**. The variable N has the **compounding distribution**.

For example, if N is Poisson, then S is **compound Poisson**.

We can find $\mathbb{E}[S]$ and $\mathrm{Var}(S)$ by conditioning on N, then using double averaging and the conditional variance formula.

Proposition 12.2 Let $S = X_1 + X_2 + \cdots + X_N$ be a compound random variable with each $X_i \sim X$. If $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[N] < \infty$, then

$$\mathbb{E}[S] = \mathbb{E}[X] \cdot \mathbb{E}[N]$$

If, in addition, X and N have finite second moments, then

$$\mathrm{Var}(S) = \mathrm{Var}(X) \cdot \mathbb{E}[N] + \mathbb{E}[X]^2 \cdot \mathrm{Var}(N).$$

Proof. For the mean, $\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S|N]]$. First,

$$\begin{aligned}
\mathbb{E}[S|N=n] &= \mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N=n\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i \,\middle|\, N=n\right] \\
&\overset{\text{independence}}{=} \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&\overset{\text{linearity}}{=} \sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= \sum_{i=1}^{n} \mathbb{E}[X].
\end{aligned}$$

since $X_i \sim X$. Hence $\mathbb{E}[S|N=n] = n \cdot \mathbb{E}[X]$, so $\mathbb{E}[S|N] = N \cdot \mathbb{E}[X]$, and therefore $\mathbb{E}[S] = \mathbb{E}[N \cdot \mathbb{E}[X]] = \mathbb{E}[X] \cdot \mathbb{E}[N]$.

For the variance, use

$$\text{Var}(S) = \mathbb{E}[\text{Var}(S|N)] + \text{Var}(\mathbb{E}[S|N]).$$

One piece at a time:

1. $\text{Var}(\mathbb{E}[S|N]) = \text{Var}(N \cdot \mathbb{E}[X]) = \mathbb{E}[X]^2 \cdot \text{Var}(N)$.
2. Also $\text{Var}(S|N) = \mathbb{E}\left[S^2|N\right] - \mathbb{E}[S|N]^2$. $\mathbb{E}\left[S^2|N\right]$ is not directly known. For $N = n$,

$$\mathbb{E}\left[S^2|N=n\right] = \mathbb{E}\left[\left(\sum_{i=1}^{N} X_i\right)^2 \,\middle|\, N=n\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right].$$

   For any random variable Q, $\mathbb{E}\left[Q^2\right] = \text{Var}(Q) + \mathbb{E}[Q]^2$, since $\mathbb{E}\left[Q^2\right]$ is the second moment. So

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] &= \text{Var}\left(\sum_{i=1}^{n} X_i\right) + \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]^2 \\
&= \sum_{i=1}^{n} \text{Var}(X_i) + \left(\sum_{i=1}^{n} \mathbb{E}[X_i]\right)^2 \\
&= n \cdot \text{Var}(X) + (n \cdot \mathbb{E}[X])^2.
\end{aligned}$$

Therefore,

$$\mathrm{Var}(S|N) = N \cdot \mathrm{Var}(X) + (N \cdot \mathbb{E}[X])^2 - (N \cdot \mathbb{E}[X])^2 = N \cdot \mathrm{Var}(X),$$

so $\mathbb{E}[\mathrm{Var}(S|N)] = \mathbb{E}[N \cdot \mathrm{Var}(X)] = \mathrm{Var}(X) \cdot \mathbb{E}[N]$.

Finally,

$$\mathrm{Var}(S) = \mathrm{Var}(X) \cdot \mathbb{E}[N] + \mathbb{E}[X]^2 \cdot \mathrm{Var}(N).$$

□

Example 12.3 The number of customers entering a store is a Poisson random variable with 20 people per day on average. The amount each customer spends is exponentially distributed with average \$100. Find the mean and variance of daily revenue.

Solution. Recall that if $X \sim \mathrm{Poisson}(\lambda)$, then $\mathbb{E}[X] = \mathrm{Var}(X) = \lambda$, while if $Y \sim \mathrm{Exp}(\lambda)$, then $\mathbb{E}[Y] = 1/\lambda$ and $\mathrm{Var}(Y) = 1/\lambda^2$. Assuming customers are independent, and spending is independent of the number of customers, let

$$S = \text{total revenue} = X_1 + \cdots + X_N,$$

where $N \sim \mathrm{Poisson}(20)$ and $X_i \sim \mathrm{Exp}(1/100)$. Then

$$\mathbb{E}[S] = \mathbb{E}[X] \cdot \mathbb{E}[N] = 100 \cdot 20 = 2000.$$

Also,

$$\begin{aligned}\mathrm{Var}(S) &= \mathrm{Var}(X) \cdot \mathbb{E}[N] + \mathbb{E}[X]^2 \cdot \mathrm{Var}(N) \\ &= \mathbb{E}[N]\left(\mathrm{Var}(X) + \mathbb{E}[X]^2\right) \\ &= 20(100^2 + 100^2) \\ &= 400{,}000.\end{aligned}$$

□

Lecture 13 — Wednesday, June 04, 2014

5. Generating Functions

Power Series and Generating Functions

We first review classical generating functions for numerical sequences, then specialize to probability generating functions.

Definition 13.1 For a sequence of real numbers $a_0, a_1, a_2, \cdots = \{a_n\}$, we define

$$A(s) = \sum_{n=0}^{\infty} a_n s^n$$

whenever the series converges. Depending on $\{a_n\}$, the series will

1. converge only when $s = 0$, or
2. converge absolutely when $s \in (-R, R)$ in which case R is the radius of convergence and $A(s)$ may not converge at the endpoints, or
3. converge absolutely for all real s.

If case (2) or (3) occurs, we call $A(s)$ the **generating function** of $\{a_n\}$. In case (1), the generating function does not exist.

We already know some generating functions:

1. If $\{a_n\} = \{1\}$, then

$$A(s) = \sum_{n=0}^{\infty} 1 \cdot s^n = \frac{1}{1-s}$$

for $|s| < 1$.

2. If $\{a_n\} = \{1/n!\}$, then
$$A(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} = e^s$$
for all $s \in \mathbb{R}$.

There is a one-to-one correspondence between sequences $\{a_n\}$ and generating functions $A(s)$. Given a function $A(s)$, we can recover the coefficients $\{a_n\}$ uniquely. One way is to use a Taylor series expansion, where
$$a_n = \frac{A^{(n)}(0)}{n!}.$$
Why?
$$A(s) = a_0 + a_1 s + a_2 s^2 + \cdots \implies A'(s) = a_1 + 2a_2 s + 3a_3 s^2 + \ldots,$$
so $A(0) = a_0$, $A'(0) = a_1$, $A''(0)/2! = a_2$, and so on.

This approach always works, but there are shortcuts. We can use known combinations of simpler generating functions with known coefficients.

Proposition 13.2 Let $A(s), B(s)$ be generating functions for sequences $\{a_n\}, \{b_n\}$, with radii of convergence R_A, R_B.

1. If $C(s) = A(s) \pm B(s)$, then
$$C(s) = \sum_{n=0}^{\infty} (a_n \pm b_n) s^n,$$
so $c_n = a_n \pm b_n$, and convergence holds for $|s| < \min\{R_A, R_B\}$.
2. Term-by-term differentiation and integration work:
$$D(s) = A'(s) = \sum_{n=1}^{\infty} n a_n s^{n-1},$$
so $d_n = (n+1)a_{n+1}$, while
$$\int_0^s A(u)\,\mathrm{d}u = \sum_{n=0}^{\infty} \frac{a_n}{n+1} s^{n+1}.$$

3. If $E(s) = A(s)B(s)$, then

$$E(s) = \sum_{n=0}^{\infty} e_n s^n \qquad \text{and} \qquad e_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

The sequence $\{e_n\}$ is the **convolution** $\{a_n\} * \{b_n\}$.

Probability Generating Functions

Definition 13.3 Let X be a discrete random variable on $\{0, 1, 2, \dots\} \cup \{\infty\}$ and let $p_n = \mathbb{P}(X = n)$ for all n. Then the **probability generating function** (**pgf**) of X is the function

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n.$$

At $s = 0$, we use the convention $0^0 = 1$.

That is, $G_X(s)$ is the generating function of the sequence of probabilities. $G_X(s)$ always converges on $(-1, 1)$ and may converge outside that interval as well. Also,

$$G_X(1) = \sum_{n=0}^{\infty} p_n \cdot 1^n = \sum_{n=0}^{\infty} \mathbb{P}(X = n) \leq 1 < \infty.$$

How do we use probability generating functions?

1. Evaluate $G_X(1)$ to test propriety. If $G_X(1) = 1$, then X is proper. If $G_X(1) = c < 1$, then $\mathbb{P}(X = \infty) = 1 - c$, so X is improper.
2. Recover the coefficients p_n from G_X, for example by Taylor expansion.
3. If X is proper, then

$$G_X'(s) = \sum_{n=1}^{\infty} n p_n s^{n-1},$$

so

$$G_X'(1-) = \sum_{n=1}^{\infty} n p_n = \sum_{n=0}^{\infty} n \cdot \mathbb{P}(X = n) = \mathbb{E}[X].$$

Both sides may be infinite; $G_X'(1-)$ denotes the left derivative.

Lecture 14 — Thursday, June 05, 2014

We can differentiate $A(s)$ term by term inside its radius of convergence. The series converges uniformly on every compact subinterval of $(-R, R)$, and

$$A'(s) = \frac{\mathrm{d}}{\mathrm{d}s} \sum a_n s^n = \sum \frac{\mathrm{d}}{\mathrm{d}s} a_n s^n.$$

Recall that the pgf of X is

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n \qquad \text{and} \qquad p_n = \mathbb{P}(X = n).$$

If we know $\mathbb{P}(X = n)$, we can try to find a closed form for $G_X(s)$. Also, $G_X(s)$ uniquely determines the distribution of X.

These are the most useful things we can do with the pgf of X:

1. Determine whether X is proper by checking if $G_X(1) = 1$.
2. Recover the p_n, although this can be tedious.
3. Obtain $\mathbb{E}[X] = G'_X(1-)$, which we already proved.
4. If $\mathbb{E}[X^2] < \infty$, find the variance of X by taking the second derivative:

$$G''_X(s) = \frac{\mathrm{d}}{\mathrm{d}s} \left(0 + \sum_{n=1}^{\infty} n \cdot p_n s^{n-1} \right) = 0 + 0 + \sum_{n=2}^{\infty} \frac{\mathrm{d}}{\mathrm{d}s} n p_n s^{n-1} = \sum_{n=2}^{\infty} n(n-1) p_n s^{n-2}.$$

Notice that

$$G''_X(1-) = \sum_{n=2}^{\infty} n(n-1) p_n = \mathbb{E}[X(X-1)],$$

which is the second **factorial moment**. In general, whenever the rth factorial moment is finite,

$$G_X^{(r)}(1-) = \mathbb{E}\left[X^{(r)}\right],$$

where

$$X^{(r)} = X(X-1)(X-2)\cdots(X-r+1)$$

is an rth degree polynomial in X.

Now,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 \\
&= \mathbb{E}\left[X^2 - X + X\right] - \mathbb{E}\left[X\right]^2 \\
&= \mathbb{E}\left[X(X-1)\right] + \mathbb{E}\left[X\right] - \mathbb{E}\left[X\right]^2 = G''_X(1-) + G'_X(1-) - (G'_X(1-))^2.
\end{aligned}$$

There is also an interesting connection to mgfs. Assuming X is proper,

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n = \sum_{n=0}^{\infty} s^n \mathbb{P}(X = n) = \mathbb{E}\left[s^X\right].$$

The moment generating function is $M_X(t) = \mathbb{E}\left[e^{tX}\right]$. Wherever the relevant expectations are finite, and with $s = e^t > 0$,

$$M_X(t) = G_X(e^t) \qquad \text{and} \qquad G_X(s) = M_X(\ln s).$$

Proposition 14.1 (Multiplicative property) If $X_1, \ldots, X_n$ are independent proper non-negative integer-valued random variables and $Y = X_1 + \cdots + X_n$, then

$$G_Y(s) = \prod_{i=1}^{n} G_{X_i}(s), \qquad 0 \le s \le 1.$$

Proof.

$$G_Y(s) = \mathbb{E}\left[s^Y\right] = \mathbb{E}\left[s^{X_1+\cdots+X_n}\right] = \mathbb{E}\left[s^{X_1}\cdots s^{X_n}\right] \overset{\text{indep}}{=} \mathbb{E}\left[s^{X_1}\right]\cdots\mathbb{E}\left[s^{X_n}\right] = G_{X_1}(s)\cdots G_{X_n}(s).$$

$\square$

This property is especially useful for compound random variables. Let $Z = X_1 + \cdots + X_N$, where the X_i are independent and identically distributed nonnegative integer-valued variables and N is a nonnegative integer-valued random variable independent of the X_i. We set $Z = 0$ when $N = 0$. Then $G_Z(s) = \mathbb{E}\left[s^Z\right] = \mathbb{E}\left[\mathbb{E}\left[s^Z \mid N\right]\right]$ and

$$\mathbb{E}\left[s^Z \mid N = n\right] = \mathbb{E}\left[s^{X_1+\cdots+X_n} \mid N = n\right] = \mathbb{E}\left[s^{X_1+\cdots+X_n}\right] = G_{X_1}(s)\cdots G_{X_n}(s) = (G_X(s))^n.$$

So for $0 \le s \le 1$, $\mathbb{E}\left[s^Z \mid N\right] = (G_X(s))^N$, and therefore $G_Z(s) = \mathbb{E}\left[(G_X(s))^N\right] = G_N(G_X(s))$.

Lecture 15 — Monday, June 09, 2014

6. Renewal Theory

Stochastic Processes

A **stochastic process** is an indexed family of random variables. Two common forms are:

1. **Discrete:** $\{X_n \mid n = 1, 2, 3, \dots\}$. For example, let X_n = outcome on the nth Bernoulli trial, or let X_n = temperature on nth day.
2. **Continuous time:** $\{X(t) \mid t \geq 0\}$. For example, $X(t)$ may be the number of claims in $(0, t)$ or the temperature at time t.

We focus on discrete processes for now.

Let λ be an event in a stochastic process. We must know, for all n, whether λ has occurred (that is, λ's occurrence must be based on $X_1, X_2, \dots, X_n$, but not $X_{n+1}, X_{n+2}, \dots$).

Definition 15.1 The **waiting time** T_λ is the number of trials until the first occurrence of λ. If $T_\lambda = n$, then we first see λ on trial n. We write $T_\lambda = \infty$ to mean that λ never occurs.

The waiting time $T_\lambda^{(k,k+1)}$ is the number of trials between the kth and $(k+1)$st occurrence of λ, given that λ has already occurred k times.

$$- - - - \lambda - - \lambda - - \cdots$$

Here $T_\lambda = 5$ and $T_\lambda^{(1,2)} = 3$, and so on.

Definition 15.2 Let $f_\lambda = \mathbb{P}(T_\lambda < \infty)$. If $f_\lambda = 1$, then T_λ is **proper** (so λ occurs eventually). If $f_\lambda < 1$, then T_λ is **improper** (so there is a chance that λ never occurs).

Definition 15.3 We say that λ is a **renewal event** when $T_\lambda, T_\lambda^{(1,2)}, T_\lambda^{(2,3)}, \ldots$ are independent and identically distributed. We say that λ is a **delayed renewal event** if and only if it satisfies all renewal event conditions except that T_λ has a different distribution (but is still independent of the inter-event waiting times).

Example 15.4

1. If $\{X_n\}$ is a sequence of Bernoulli(p) trials, then the event $\lambda = \mathrm{S}$ is renewal since $T_\lambda, T_\lambda^{(k,k+1)} \sim \mathrm{Geo}(p)$.

   The event $\lambda = \mathrm{SF}$ (that is, λ occurs on n if trial n was F and trial $n-1$ was S) is also renewal:

   $$\mathrm{F\ S\ F\ S\ S\ F\ F\ F\ldots S\ F.}$$

   Because the pattern has no nonempty proper suffix that is also a prefix, an occurrence leaves no partial copy of the next occurrence in place. The unused Bernoulli trials are independent of the past, so the inter-occurrence times are independent and have the same distribution as T_λ.

   What if $\lambda = \mathrm{SS}$? Then

   $$\mathrm{F\ S\ F\ (S\ S)\ S.}$$

   We have $\mathbb{P}(T_\lambda = 2) = p^2$ and $\mathbb{P}(T_\lambda = 1) = 0$. But $\mathbb{P}(T_\lambda^{(1,2)} = 1) = p$ and $\mathbb{P}(T_\lambda^{(1,2)} = 2) = qp$. Hence $T_\lambda \neq T_\lambda^{(1,2)}$, so λ is not renewal. It is a delayed renewal event because of overlap.

2. Roll a fair six-sided die repeatedly, and let X_n be the number shown on the nth roll.

   (a) Let $\lambda =$ "1 3 3 1".

   $$- - - - 1\ 3\ 3\ 1 \mid 3\ 3\ 1 - -$$

   $\mathbb{P}(T_\lambda = 3) = 0$, but $\mathbb{P}(T_\lambda^{(1,2)} = 3) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$, so this is delayed renewal.

   (b) $\lambda =$ "1 3 1 3" is also delayed renewal. Why? $\mathbb{P}(T_\lambda = 2) = 0$, but $\mathbb{P}(T_\lambda^{(1,2)} = 2) = \frac{1}{6} \cdot \frac{1}{6}$.

   (c) $\lambda =$ "1 1 3 3":

   $$- - 1\ 1\ 3\ 3 \mid 1\ 1\ 3\ 3$$

   This pattern has no nonempty proper suffix that is also a prefix. After each occurrence, the unused rolls are independent of the past and the pattern search restarts from its initial state. Therefore

$$\mathbb{P}(T_\lambda = n) = \mathbb{P}(T_\lambda^{(1,2)} = n)$$

for all n, so λ is renewal.

Lecture 16 — Wednesday, June 11, 2014

Recurrent and Transient Events

Recall that λ is an event in a stochastic process. We defined T_λ as the waiting time until the first occurrence of λ, $f_\lambda = \mathbb{P}(T_\lambda < \infty)$, and $T_\lambda^{(k,k+1)}$ as the inter-event waiting time between the kth and $(k+1)$st occurrence of λ. We said that λ is renewal when T_λ and all $T_\lambda^{(k,k+1)}$ are independent and identically distributed, and λ is delayed renewal when T_λ has a different distribution.

By definition, f_λ is the probability that λ occurs at all. Thus T_λ is proper exactly when $f_\lambda = 1$.

Definition 16.1 For a renewal event λ, we say λ is **recurrent** when $f_\lambda = 1$ and **transient** when $f_\lambda < 1$.

Example 16.2 Consider Bernoulli trials with $0 < p < 1$. $\lambda =$ "S F" is renewal from last time. Is it recurrent or transient?

$$\underbrace{F \dots F\ S}_{T_S}\ \underbrace{F}_{T_F}\ \underbrace{S}_{T_S} \cdots \underbrace{S}_{T_F}\ \underbrace{F}$$

So

$$f_\lambda = \mathbb{P}(T_\lambda < \infty) = \mathbb{P}(T_S + T_F < \infty) = \mathbb{P}(T_S < \infty) \cdot \mathbb{P}(T_F < \infty) = 1 \cdot 1 = 1.$$

Hence λ is recurrent. More generally, any finite pattern composed only of symbols having positive probability occurs in a finite number of independent trials with probability 1.

Example 16.3 Roll a six-sided die. λ occurs on trial n if

$$X_1 = X_2 = \cdots = X_n = 3.$$

Is λ renewal? We have $\mathbb{P}(T_\lambda = 1) = 1/6$, and λ never occurs if $X_1 \neq 3$, so $\mathbb{P}(T_\lambda = \infty) = 5/6$. Also $\mathbb{P}(T_\lambda^{(1,2)} = 1) = 1/6$ and $\mathbb{P}(T_\lambda^{(1,2)} = \infty) = 5/6$, so λ is indeed a renewal event. It is transient because

$$f_\lambda = \mathbb{P}(T_\lambda < \infty) = \frac{1}{6} < 1.$$

Definition 16.4 We say a recurrent renewal event λ is **positive recurrent** when $\mathbb{E}[T_\lambda] < \infty$, and **null recurrent** when $\mathbb{E}[T_\lambda] = \infty$.

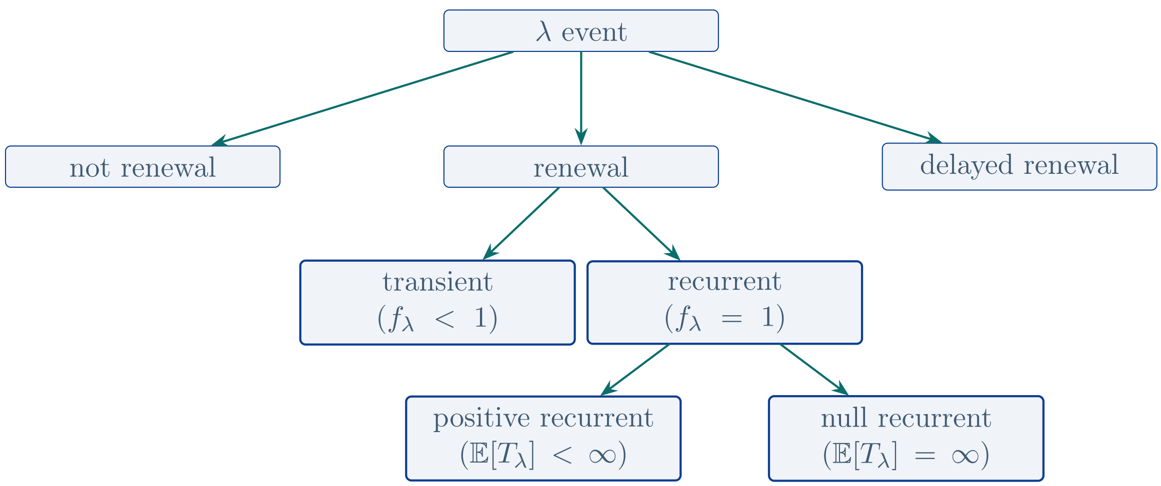

FIGURE 2

Figure 2 summarizes the renewal classification.

Definition 16.5 For a renewal event λ, let V_λ be the number of times that λ occurs over the infinite time horizon. Then $\text{Range}(V_\lambda) = \{0, 1, 2, \dots\} \cup \{\infty\}$.

Proposition 16.6 The pmf of V_λ is

$$\mathbb{P}(V_\lambda = k) = \begin{cases} (f_\lambda)^k (1 - f_\lambda), & k = 0, 1, 2, \dots, \\ 0, & k = \infty \text{ and } f_\lambda < 1, \\ 1, & k = \infty \text{ and } f_\lambda = 1. \end{cases}$$

Proof. For $k \neq \infty$:

$$\mathbb{P}(V_\lambda = k) = \mathbb{P}(\lambda \text{ occurs } k \text{ times})$$

$$
\begin{aligned}
&= \mathbb{P}(T_\lambda < \infty, T_\lambda^{(1,2)} < \infty, \ldots, T_\lambda^{(k-1,k)} < \infty, T_\lambda^{(k,k+1)} = \infty) \\
&= \mathbb{P}(T_\lambda < \infty) \cdot \mathbb{P}(T_\lambda^{(1,2)} < \infty) \ldots \mathbb{P}(T_\lambda^{(k,k+1)} = \infty) \\
&= f_\lambda \cdot f_\lambda \ldots (1 - f_\lambda) = (f_\lambda)^k (1 - f_\lambda).
\end{aligned}
$$

For infinitely many occurrences, continuity from above gives

$$\mathbb{P}(V_\lambda = \infty) = \lim_{k\to\infty} \mathbb{P}(V_\lambda \geq k) = \lim_{k\to\infty} (f_\lambda)^k,$$

which is zero when $f_\lambda < 1$ and one when $f_\lambda = 1$. $\square$

Lecture 17 — Thursday, June 12, 2014

The Renewal Relation and the Renewal Sequence

Recall that a renewal event λ is recurrent when

$$f_\lambda = 1 = \mathbb{P}(T_\lambda < \infty),$$

and transient when $f_\lambda < 1$. Recurrent events occur infinitely often, while transient events cannot occur infinitely often. The average number of occurrences is

$$\mathbb{E}[V_\lambda] = \begin{cases} \infty, & \lambda \text{ recurrent}, \\ \dfrac{f_\lambda}{1 - f_\lambda}, & \lambda \text{ transient}. \end{cases}$$

Why is this so? If $f_\lambda < 1$, then $\mathbb{P}(V_\lambda = k) = (f_\lambda)^k(1 - f_\lambda)$. Here $V_\lambda + 1 \sim \text{Geo}(1 - f_\lambda)$, and

$$\mathbb{E}[\text{Geo}(1 - f_\lambda)] = \frac{1}{1 - f_\lambda}.$$

Therefore $\mathbb{E}[V_\lambda + 1] = \dfrac{1}{1 - f_\lambda}$, and hence $\mathbb{E}[V_\lambda] = \dfrac{f_\lambda}{1 - f_\lambda}$.

If we know f_λ, we can find $\mathbb{E}[V_\lambda]$ and determine whether λ is transient or recurrent. However, f_λ can be hard to obtain, and does not tell us whether λ is positive or null recurrent. For that, we use the renewal sequence.

Definition 17.1 For a renewal event λ, the **renewal sequence** $\{r_n\}$ is defined by $r_0 = 1$ and $r_n = \mathbb{P}(\lambda \text{ occurs on trial } n)$.

Warning $r_n \neq \mathbb{P}(\lambda \text{ occurs on trial } n \text{ for the first time})$.

We use this sequence in two ways:

1. Find $\mathbb{E}[V_\lambda]$ (which determines recurrence versus transience). Write

$$V_\lambda = \sum_{n=1}^{\infty} \mathbb{1}_n \qquad \text{and} \qquad \mathbb{1}_n = \begin{cases} 1, & \lambda \text{ occurs on } n, \\ 0, & \text{otherwise.} \end{cases}$$

   Then

$$\mathbb{E}[V_\lambda] = \sum_{n=1}^{\infty} \mathbb{E}[\mathbb{1}_n] = \sum_{n=1}^{\infty} \mathbb{P}(\lambda \text{ occurs on trial } n) = \sum_{n=1}^{\infty} r_n.$$

   If

$$\sum_{n=1}^{\infty} r_n = \infty,$$

   then λ is recurrent and $f_\lambda = 1$. If

$$\sum_{n=1}^{\infty} r_n = \mathbb{E}[V_\lambda] < \infty,$$

   then λ is transient and

$$\mathbb{E}[V_\lambda] = \frac{f_\lambda}{1 - f_\lambda} \implies f_\lambda = \frac{\mathbb{E}[V_\lambda]}{1 + \mathbb{E}[V_\lambda]}.$$

2. Obtain information about T_λ, including propriety, short proper versus null proper, $\mathbb{E}[T_\lambda]$, $\text{Var}(T_\lambda)$, and $\mathbb{P}(T_\lambda = n)$. This information is encoded in the pgf of T_λ, denoted $F_\lambda(s)$. Set $f_0 = 0$,

$$f_n = \mathbb{P}(T_\lambda = n) = \mathbb{P}(\lambda \text{ first occurs on trial } n).$$

   Then

$$F_\lambda(s) = \sum_{n=0}^{\infty} f_n s^n,$$

   the pgf of T_λ. We set $f_0 = 0$ because $T_\lambda = 0$ is impossible, so $F_\lambda(s)$ has no constant term.

Let

$$R_\lambda(s) = \sum_{n=0}^{\infty} r_n s^n$$

be the generating function for $\{r_n\}$.

Theorem 17.2 (Renewal relation) For a renewal event λ, as a formal power-series identity and hence for $0 \le s < 1$,

$$F_\lambda(s) = 1 - \frac{1}{R_\lambda(s)}.$$

Proof. For $n \ge 1$,

$$\begin{aligned}
r_n = \mathbb{P}(\lambda \text{ occurs on trial } n) &= \sum_{k=1}^{n} \mathbb{P}(\lambda \text{ first occurs on } k \text{ and also on } n) \\
&= \sum_{k=1}^{n} \mathbb{P}(\lambda \text{ first occurs on } k) \cdot \mathbb{P}(\lambda \text{ occurs on } n \mid \lambda \text{ occurs on } k) \\
&= \sum_{k=1}^{n} f_k \cdot r_{n-k} \\
&= \sum_{k=0}^{n} f_k \cdot r_{n-k} \quad \text{since } f_0 = 0.
\end{aligned}$$

So $\{r_n\}_{n\ge1}$ is the convolution of $\{f_n\}$ and $\{r_n\}$. Now,

$$R_\lambda(s) = \sum_{n=0}^{\infty} r_n s^n = 1 + \sum_{n=1}^{\infty} r_n s^n = 1 + \sum_{n=1}^{\infty}\left(\sum_{k=0}^{n} f_k r_{n-k}\right) s^n = 1 + \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} f_k r_{n-k}\right) s^n,$$

since $f_0 = 0$. That sum is the generating function of the convolution, so it equals the product of the generating functions of the two sequences. Therefore,

$$R_\lambda(s) = 1 + F_\lambda(s)\, R_\lambda(s).$$

Solving gives $F_\lambda(s) = 1 - \dfrac{1}{R_\lambda(s)}$. $\square$

Lecture 18 — Monday, June 16, 2014

Applications of Renewal Sequences

Recall that the renewal sequence is $\{r_n\}$ with $r_0 = 1$ and $r_n = \mathbb{P}(\lambda \text{ occurs on } n)$. We found that

$$\mathbb{E}[V_\lambda] = \sum_{n=1}^{\infty} r_n.$$

Example 18.1

1. Consider Bernoulli trials and $\lambda =$ "S F". Then $r_0 = 1$, $r_1 = 0$, $r_2 = p(1-p)$, $r_3 = p(1-p)$, $r_4 = p(1-p), \ldots, r_n = p(1-p), \ldots$. So

$$\mathbb{E}[V_\lambda] = \sum_{n=1}^{\infty} r_n = 0 + \sum_{n=2}^{\infty} p(1-p) = \infty$$

for $0 < p < 1$. Hence λ is recurrent.

2. Consider rolling a fair 6-sided die and let $\lambda =$ all rolls up to and including roll n were "3". Then $r_0 = 1$, $r_1 = 1/6$, $r_2 = (1/6)^2, \ldots, r_n = (1/6)^n$. So

$$\mathbb{E}[V_\lambda] = \sum_{n=1}^{\infty} r_n = \sum_{n=1}^{\infty} (1/6)^n = \frac{1/6}{1-1/6} = \frac{1}{5} < \infty.$$

Thus λ is transient, and

$$f_\lambda = \frac{\mathbb{E}[V_\lambda]}{1+\mathbb{E}[V_\lambda]} = \frac{1/5}{1+1/5} = \frac{1}{6}.$$

The other use of the renewal sequence is the renewal relation:

$$F_\lambda(s) = 1 - \frac{1}{R_\lambda(s)},$$

where $F_\lambda(s)$ is the pgf of T_λ and $R_\lambda(s)$ is the generating function of $\{r_n\}$.

Example 18.2 Roll a fair 6-sided die. Let $\lambda =$ "2 4", which is a renewal event. Classify the propriety and recurrence of λ, and find the probability that it first occurs on trial n, for $n = 0, \ldots, 4$.

Solution. We have $r_0 = 1$, $r_1 = 0$, $r_2 = 1/36$, $r_3 = 1/36, \ldots, r_n = 1/36$ for $n \geq 2$. So

$$R_\lambda(s) = \sum_{n=0}^{\infty} r_n s^n = 1 + \frac{1}{36}s^2 + \frac{1}{36}s^3 + \cdots = 1 + \frac{1}{36}s^2(1 + s + s^2 + \ldots) = 1 + \frac{s^2/36}{1-s}.$$

Then

$$F_\lambda(s) = 1 - \frac{1}{R_\lambda(s)} = 1 - \frac{1}{1 + \frac{s^2/36}{1-s}} = 1 - \frac{1-s}{1-s+s^2/36} = \frac{s^2/36}{1-s+s^2/36}.$$

Now $F_\lambda(1) = 1$, so T_λ is proper and λ is recurrent.

We can compute $F'_\lambda(s)$, and $F'_\lambda(1-) = 36$, so

$$\mathbb{E}\left[T_\lambda\right] = F'_\lambda(1-) = 36 < \infty,$$

hence T_λ is short proper and λ is positive recurrent. We can find $\mathrm{Var}(T_\lambda)$ using $F''_\lambda(s)$ (or higher moments), and we can find $\mathbb{P}(T_\lambda = n) = f_n$. Write

$$F_\lambda(s) = \left(\frac{s^2}{36}\right)\left(1 - s + \frac{s^2}{36}\right)^{-1}.$$

Expand the second factor using Taylor series. If

$$A(0) = 1, \quad A'(0) = 1, \quad A''(0) = \frac{35}{18}, \ \ldots,$$

then

$$F_\lambda(s) = \frac{s^2}{36}\left(\frac{1}{0!}s^0 + \frac{1}{1!}s^1 + \frac{35}{18}\frac{1}{2!}s^2 + \ldots\right) = \frac{1}{36}s^2 + \frac{1}{36}s^3 + \frac{35}{36^2}s^4 + \ldots.$$

So

$$f_0 = 0, f_1 = 0, f_2 = f_3 = 1/36, f_4 = \frac{35}{36^2},$$

and so on. These values are consistent with direct counting: $\mathbb{P}(T_\lambda = 1) = 0$; for f_2 and f_3, we need a 2 and a 4 to occur; for f_4, pattern "24" must occur on trial 4 but not earlier, giving ??24 except ?? = 24, so 35 possibilities. Thus

$$\left(\frac{1}{6}\right)^2 - \left(\frac{1}{6}\right)^4 = \frac{35}{36^2} = f_4$$

. $\square$

In summary:

$$
\begin{aligned}
F_\lambda(1) = 1 &\implies \lambda \text{ is recurrent} \\
F_\lambda(1) < 1 &\implies \lambda \text{ is transient} \\
F_\lambda(1) = 1 \text{ and } F'_\lambda(1-) = \mathbb{E}\left[T_\lambda\right] < \infty &\implies \lambda \text{ is positive recurrent,} \\
F_\lambda(1) = 1 \text{ and } F'_\lambda(1-) = \infty &\implies \lambda \text{ is null recurrent.}
\end{aligned}
$$

When T_λ is proper and has a finite second moment, the second derivative gives the second factorial moment, so

$$\operatorname{Var}(T_\lambda) = F''_\lambda(1-) + F'_\lambda(1-) - F'_\lambda(1-)^2.$$

The probabilities $f_n = \mathbb{P}(T_\lambda = n)$ come from the Taylor expansion.

Delayed Renewal Events

Definition 18.3 λ is **delayed renewal** if $T_\lambda, T_\lambda^{(1,2)}, T_\lambda^{(2,3)}, \ldots$ are independent, and all waiting times after T_λ are identically distributed.

Let $\bar{\lambda}$ be the event "λ given that λ has already occurred". Then

$$T_{\bar{\lambda}} = T_\lambda^{(1,2)}, T_{\bar{\lambda}}^{(1,2)} = T_\lambda^{(2,3)}, \ldots, T_{\bar{\lambda}}^{(k,k+1)} = T_\lambda^{(k+1,k+2)}, \ldots$$

are independent and identically distributed, so $\bar{\lambda}$ is renewal.

Definition 18.4 For a delayed renewal event λ, the **delayed renewal sequence** $\{d_n\}$ is defined by $d_0 = 0$ and $d_n = \mathbb{P}(\lambda \text{ occurs on trial } n)$ for $n \geq 1$. The associated renewal sequence (that is, the renewal sequence for the associated renewal event $\bar{\lambda}$) has $\bar{r}_0 = 1$ and, for $n \geq 1$,

$$
\begin{aligned}
\bar{r}_n &= \mathbb{P}(\bar{\lambda} \text{ occurs on trial } n) \\
&= \mathbb{P}(\lambda \text{ occurs on trial } n \mid \text{we just saw } \lambda \text{ on trial } 0).
\end{aligned}
$$

So $\bar{r}_0 = 1$ means that we just saw λ.

Lecture 19 — Wednesday, June 18, 2014

Recall that for a delayed renewal event λ, the delayed renewal sequence is

$$d_0 = 0 \qquad \text{and} \qquad d_n = \mathbb{P}(\lambda \text{ occurs on trial } n).$$

Its generating function is

$$D_\lambda(s) = \sum_{n=0}^{\infty} d_n s^n.$$

The associated renewal event $\bar{\lambda}$ is conditional on λ having already occurred. Its renewal sequence is

$$\bar{r}_0 = 1 \qquad \text{and} \qquad \bar{r}_n = \mathbb{P}(\bar{\lambda} \text{ occurs on trial } n).$$

Why do $r_0 = \bar{r}_0 = 1$ but $d_0 = 0$? If λ is renewal, then whenever it occurs the process restarts, so while waiting for λ the first time we may as well assume we just observed λ, giving $r_0 = 1$. The same applies to $\bar{\lambda}$ in the delayed case. But for delayed λ, we cannot assume we just observed λ at time 0, so $d_0 = 0$.

Now, let

$$R_{\bar{\lambda}}(s) = \sum_{n=0}^{\infty} \bar{r}_n s^n.$$

Then

$$F_\lambda(s) = \frac{D_\lambda(s)}{R_{\bar{\lambda}}(s)},$$

which gives information about T_λ. We can also obtain information about $T_\lambda^{(k,k+1)}$, which is just $T_{\bar{\lambda}}$. Hence

$$F_{\bar{\lambda}}(s) = 1 - \frac{1}{R_{\bar{\lambda}}(s)}.$$

Thus, we have shown the following:

Theorem 19.1 (Delayed renewal relation) If λ is delayed renewal and $\bar{\lambda}$ is its associated renewal event, then

$$F_\lambda(s) = \frac{D_\lambda(s)}{R_{\bar{\lambda}}(s)} \qquad \text{and} \qquad F_{\bar{\lambda}}(s) = 1 - \frac{1}{R_{\bar{\lambda}}(s)}.$$

Proof. Let $f_k = \mathbb{P}(T_\lambda = k)$. For $n \geq 1$, partition according to the time k of the first occurrence. Once that occurrence happens, the delayed process restarts according to the associated renewal sequence, so

$$d_n = \sum_{k=1}^{n} f_k \bar{r}_{n-k}.$$

This convolution and the facts $d_0 = f_0 = 0$, $\bar{r}_0 = 1$ give

$$D_\lambda(s) = F_\lambda(s) R_{\bar{\lambda}}(s).$$

Since $R_{\bar{\lambda}}(s)$ has constant term 1, it is invertible as a formal power series, and hence

$$F_\lambda(s) = \frac{D_\lambda(s)}{R_{\bar{\lambda}}(s)}.$$

Applying Theorem 17.2 to the associated renewal event $\bar{\lambda}$ gives the second identity. □

Example 19.2 Roll a six-sided die, with λ = "2 2" delayed. Then $d_0 = 0$, $d_1 = \mathbb{P}(\lambda \text{ occurs on trial } 1) = 0$, $d_2 = (1/6)^2$, $d_3 = (1/6)^2$, ... So

$$D_\lambda(s) = \sum_{n=0}^{\infty} d_n s^n = 0 + 0 + \sum_{n=2}^{\infty} (1/6)^2 s^n = \frac{1}{36} s^2 \sum_{n=0}^{\infty} s^n = \frac{s^2/36}{1-s}.$$

For the associated renewal event $\bar{\lambda}$ = "22" | "22", we have $\bar{r}_0 = 1$, $\bar{r}_1 = 1/6$ by overlap, $\bar{r}_2 = 1/36$, $\bar{r}_3 = 1/36$, ... Thus

$$R_{\bar{\lambda}}(s) = 1 + \frac{s}{6} + \frac{s^2/36}{1-s} = \frac{(1-s) + \frac{s}{6}(1-s) + s^2/36}{1-s}.$$

Then

$$F_\lambda(s) = \frac{D_\lambda(s)}{R_{\bar{\lambda}}(s)} = \frac{s^2/36}{(1-s) + \frac{s}{6}(1-s) + s^2/36}.$$

Check $F_\lambda(1) = 1$. So T_λ is proper and λ is guaranteed to occur. Also,

$$\mathbb{E}\left[T_\lambda\right] = F'_\lambda(1-) = 42 \neq 36.$$

We can similarly find

$$F_{\bar{\lambda}}(s) = 1 - \frac{1-s}{(1-s) + \frac{s}{6}(1-s) + s^2/36},$$

the pgf of inter-event waiting times. Since $F_{\bar{\lambda}}(1) = 1$, $\bar{\lambda}$ is recurrent. Therefore: (1) λ is guaranteed to occur once, and (2) once it occurs, it will occur infinitely often. So λ occurs infinitely often. Also $\mathbb{E}[T_{\bar{\lambda}}] = F'_{\bar{\lambda}}(1-) = 36$. This is not the same distribution as T_λ for $\lambda =$ "2 4", although it has the same mean.

Lecture 20 — Thursday, June 19, 2014

Renewal Theorems and Random Walks

Definition 20.1 Let λ be a renewal event with renewal sequence $\{r_n\}$, and let

$$d = \gcd\{n \mid r_n > 0\}.$$

If $d = 1$, then λ is **aperiodic**. Otherwise, λ is **periodic** with **period** d.

Theorem 20.2 (Aperiodic renewal theorem) Suppose λ is renewal and aperiodic. Then

$$\lim_{n\to\infty} r_n = \frac{1}{\mathbb{E}[T_\lambda]},$$

with the convention $1/\infty = 0$.

If an event can occur only on multiples of d, then $\lim_{n\to\infty} r_n$ does not exist; in that case we take limits along the subsequence $n = kd$:

Theorem 20.3 (Periodic renewal theorem) If λ is renewal and periodic with period $d > 1$, then

$$\lim_{n\to\infty} r_{nd} = \frac{d}{\mathbb{E}[T_\lambda]},$$

again with $d/\infty = 0$.

Definition 20.4 A **random walk** is a discrete stochastic process. Its states lie in

$\{\dots,-2,-1,0,1,2,\dots\}$, and it is defined by

$$X_0 = 0 \qquad \text{and} \qquad X_i = \begin{cases} X_{i-1}+1 & \text{with probability } p, \\ X_{i-1}-1 & \text{with probability } 1-p. \end{cases}$$

where $0 < p < 1$ and $q = 1-p$.

Let λ_{00} = "return to 0". This is a renewal event, since the process restarts whenever it reaches 0. We observe that λ_{00} can occur only on even numbers of trials, so $d = 2$ and the event is periodic. Thus $r_{2n+1} = 0$ for all $n \in \mathbb{N}$ and $r_0 = 1$. Also,

$$r_{2n} = \mathbb{P}(\text{back to 0 on trial } 2n) = \mathbb{P}(n \text{ up steps and } n \text{ down steps}) = \binom{2n}{n} p^n q^n.$$

We use the renewal relation to obtain the pgf of T_{00}. We need three lemmas:

1. The **generalized binomial theorem** states that, for $a \in \mathbb{R}$ and $|x| < 1$,
$$(1+x)^a = \sum_{n=0}^{\infty} \binom{a}{n} x^n.$$
2. An induction argument shows that
$$\binom{2n}{n} = \binom{-1/2}{n}(-4)^n.$$
3. If $0 < p < 1$ and $q = 1-p$, then
$$1 - 4pq = (p-q)^2.$$

We want $F_{00}(s)$, the pgf of T_{00}. We compute

$$R_{00}(s) = \sum_{n=0}^{\infty} r_n s^n = \sum_{n=0}^{\infty} r_{2n} s^{2n} = \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n s^{2n} = \sum_{n=0}^{\infty} \binom{-1/2}{n} (-4pqs^2)^n = (1-4pqs^2)^{-1/2}.$$

Hence
$$F_{00}(s) = 1 - \frac{1}{R_{00}(s)} = 1 - \sqrt{1-4pqs^2}.$$

Now

$$f_{00} = \mathbb{P}(\lambda_{00} \text{ occurs}) = F_{00}(1) = 1 - \sqrt{1-4pq} = 1 - \sqrt{(p-q)^2} = 1 - |p-q|$$
$$= \begin{cases} 1 & \text{if } p = q = 1/2, \\ 1 - |p-q| & \text{if } p \neq q. \end{cases}$$

Lecture 21 — Monday, June 23, 2014

Recall that for a random walk, λ_{00} = "return to 0". We derived $F_{00}(s) = 1 - \sqrt{1-4pqs^2}$ and found

$$f_{00} = \mathbb{P}(T_{00} < \infty) = F_{00}(1) = \begin{cases} 1, & p = 1/2 \\ 1 - |p-q|, & p \neq 1/2 \end{cases}.$$

If $p = 1/2$ (i.e., we have a **balanced random walk**), λ_{00} is recurrent and occurs infinitely often. If $p \neq 1/2$, λ_{00} is transient and occurs only finitely many times. If $p = 1/2$, then $\mathbb{E}[V_{00}] = \infty$, while if $p \neq 1/2$,

$$\mathbb{E}[V_{00}] = \frac{f_{00}}{1-f_{00}} = \frac{1-|p-q|}{|p-q|}.$$

For example, if $p = 0.55$, then $f_{00} = 0.9$ and $\mathbb{E}[V_{00}] = 0.9/0.1 = 9$. If $p = 0.7$, then $f_{00} = 0.6$ and $\mathbb{E}[V_{00}] = 0.6/0.4 = 1.5$.

Also,

$$F'_{00}(s) = -\frac{1}{2}(1-4pqs^2)^{-1/2}(-8pqs) = \frac{4pqs}{\sqrt{1-4pqs^2}}.$$

At $p = q = 1/2$, this is $s/\sqrt{1-s^2}$, which goes to ∞ as $s \to 1$. Therefore λ_{00} is null recurrent when $p = 1/2$. If $p \neq 1/2$, then λ_{00} is transient, so $\mathbb{E}[T_{00}] = \infty$.

For a random walk, let λ_{01} = "get to state 1, starting from 0".

0	-1	-2	-1	0	1	2	1	2	3	2	1	0	-1	0	1
					5		2				4				4

FIGURE 3

Figure 3 shows that λ_{01} is a delayed renewal event. The associated event $\bar{\lambda}_{01}$ is the same as λ_{11} = "return to 1", which behaves the same way as λ_{00}. Hence $R_{\bar{\lambda}_{01}}(s) = R_{00}(s)$. Also, $d_0 = 0$,

$d_1 = p$, $d_{2n} = 0$ for all n, and

$$d_{2n+1} = \binom{2n+1}{n} p^{n+1} q^n$$

since we need $n+1$ up steps and n down steps. Therefore

$$D_{01}(s) = \sum_{n=0}^{\infty} d_n s^n = \sum_{n=0}^{\infty} d_{2n+1} s^{2n+1} = \sum_{n=0}^{\infty} \binom{2n+1}{n} p^{n+1} q^n s^{2n+1}.$$

For the derivation below assume $0 < p < 1$; the case $p = 1$, if admitted as a degenerate walk, has $T_{01} = 1$ directly. Now

$$\binom{2n+1}{n+1} = \frac{(2n+1)!}{(n+1)!n!}$$

and

$$\binom{2n+2}{n+1} = \frac{(2n+2)(2n+1)!}{(n+1)!(n+1)!} = 2\binom{2n+1}{n+1}.$$

So

$$\begin{aligned}
D_{01}(s) &= \sum_{n=0}^{\infty} \frac{1}{2}\binom{2n+2}{n+1} p^{n+1} q^n s^{2n+1} = \sum_{n=0}^{\infty} \binom{-1/2}{n+1} (-4)^{n+1} \frac{1}{2} p^{n+1} q^n s^{2n+1} \\
&= \frac{1}{2qs} \sum_{n=0}^{\infty} \binom{-1/2}{n+1} (-4pqs^2)^{n+1} = \frac{1}{2qs} \sum_{k=1}^{\infty} \binom{-1/2}{k} (-4pqs^2)^k \\
&= \frac{1}{2qs}\left((1-4pqs^2)^{-1/2} - 1\right).
\end{aligned}$$

Finally,

$$F_{01}(s) = \frac{D_{01}(s)}{R_{00}(s)} = \frac{1}{2qs}\left((1-4pqs^2)^{-1/2} - 1\right) \cdot \sqrt{1-4pqs^2} = \frac{1-\sqrt{1-4pqs^2}}{2qs}.$$

Hence

$$f_{01} = F_{01}(1) = \mathbb{P}(\lambda_{01} \text{ occurs}) = \frac{1-|p-q|}{2q} = \begin{cases} 1, & p \geq q \\ p/q, & p < q \end{cases}.$$

So if $p \geq 1/2$, λ_{01} is guaranteed to occur; if $p < 1/2$, it may fail to occur. If $p \geq 1/2$, we can also find

$$\mathbb{E}\left[T_{01}\right] = F_{01}'(1-) = \begin{cases} \dfrac{1}{p-q}, & \text{if } p > 1/2 \\ \infty, & \text{if } p = 1/2 \end{cases},$$

and $\mathbb{E}\left[T_{01}\right] = \infty$ if $p < 1/2$. Thus T_{01} is short proper if $p > 1/2$, null proper if $p = 1/2$, and improper if $p < 1/2$.

Now consider T_{0k}, the waiting time to reach state k. If $k > 0$, then

$$T_{0k} = T_{01} + T_{12} + \cdots + T_{k-1,k},$$

The argument uses the standard restart principle: after the walk first reaches a state, its future steps are independent of the steps already taken and have the same law as those of a fresh walk from that state. A complete proof of this principle at random hitting times requires the strong Markov property, which is beyond the scope of this course. Accepting the restart principle, spatial homogeneity shows that each $T_{j,j+1}$ has the same distribution as T_{01} and that the successive passage times are independent. Therefore

$$\mathbb{E}\left[T_{0k}\right] = k \cdot \mathbb{E}\left[T_{01}\right] = \begin{cases} \dfrac{k}{p-q}, & p > 1/2 \\ \infty, & p \leq 1/2 \end{cases},$$

and

$$f_{0k} = \mathbb{P}(T_{0k} < \infty) = \mathbb{P}(T_{01} < \infty, T_{12} < \infty, \dots) = (f_{01})^k = \begin{cases} 1, & p \geq 1/2 \\ (p/q)^k, & p < 1/2 \end{cases}.$$

Lecture 22 — Wednesday, June 25, 2014

Gambler's Ruin and Walk Variations

For the regular random walk, we know that λ_{00} is transient ($p \neq 1/2$) or null recurrent ($p = 1/2$), and λ_{01} is delayed: short ($p > 1/2$), null ($p = 1/2$), or improper ($p < 1/2$). The event λ_{0k} is equivalent to λ_{01} repeated k times, so

$$\mathbb{E}\left[T_{0k}\right] = \frac{k}{p-q}$$

if $p > 1/2$, and

$$f_{0k} = (f_{01})^k = \left(\frac{p}{q}\right)^k$$

if $p < 1/2$. If $k < 0$, we reverse the roles of p and q and proceed in the same way.

Example 22.1 (Gambler's ruin) Let $X_0 = i$. The process moves like a random walk, but with two absorbing barriers at 0 and k (so we cannot leave state 0 or k). The interpretation is that a gambler starts with wealth \$$i$, then wins or loses \$1 in each independent game. The value X_n is the gambler's wealth after n games. If $X_n = 0$, the gambler is bankrupt and stops; if $X_n = k$, play also stops.

Let

$$\mathbb{P}_i = \mathbb{P}(\text{gets to } k \text{ before going bankrupt} \mid X_0 = i).$$

Clearly, $\mathbb{P}_0 = 0$ and $\mathbb{P}_k = 1$. The quantity $\mathbb{P}_i$ increases with i and depends on p and q. Conditioning on the first game gives

$$\begin{aligned}\mathbb{P}_i &= \mathbb{P}(\text{get to } k \mid \text{win}) \cdot \mathbb{P}(\text{win}) + \mathbb{P}(\text{get to } k \mid \text{lose}) \cdot \mathbb{P}(\text{lose}) \\ &= \mathbb{P}_{i+1} \cdot p + \mathbb{P}_{i-1} \cdot q.\end{aligned}$$

Hence

$$\mathbb{P}_i \cdot p + \mathbb{P}_i \cdot q = \mathbb{P}_{i+1} \cdot p + \mathbb{P}_{i-1} \cdot q \implies (\mathbb{P}_{i+1} - \mathbb{P}_i) = \frac{q}{p}(\mathbb{P}_i - \mathbb{P}_{i-1}),$$

for $i = 1, \ldots, k-1$. This difference equation has solution

$$\mathbb{P}_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^k}, & p \neq \dfrac{1}{2} \\ \frac{i}{k}, & p = \dfrac{1}{2} \end{cases},$$

since for $p \neq q$ we can verify both boundary conditions and the recursion directly. We have

$$\mathbb{P}_0 = \frac{1-(q/p)^0}{1-(q/p)^k} = 0 \qquad \text{and} \qquad \mathbb{P}_k = \frac{1-(q/p)^k}{1-(q/p)^k} = 1.$$

Also, for $1 \leq i \leq k-1$,

$$\begin{aligned} p\mathbb{P}_{i+1} + q\mathbb{P}_{i-1} &= p\,\frac{1-(q/p)^{i+1}}{1-(q/p)^k} + q\,\frac{1-(q/p)^{i-1}}{1-(q/p)^k} \\ &= \frac{p+q-\big(p(q/p)^{i+1} + q(q/p)^{i-1}\big)}{1-(q/p)^k} \\ &= \frac{1-\big(q^{i+1}p^{-i} + q^i p^{1-i}\big)}{1-(q/p)^k} \\ &= \frac{1-\big(q^i p^{-i}(q+p)\big)}{1-(q/p)^k} = \frac{1-(q/p)^i}{1-(q/p)^k} = \mathbb{P}_i. \end{aligned}$$

When $p = q = 1/2$, the recursion becomes

$$\mathbb{P}_i = \frac{1}{2}\mathbb{P}_{i+1} + \frac{1}{2}\mathbb{P}_{i-1},$$

whose solutions are linear in i; imposing $\mathbb{P}_0 = 0$ and $\mathbb{P}_k = 1$ gives $\mathbb{P}_i = i/k$.

Example 22.2 Alice and Bob play repeated games. In each game, the loser pays the winner \$5. Bob starts with \$50 and Alice starts with \$25.

1. If each game is fair, find the probability that Bob loses all of his money.
2. If Alice wins each game with probability 0.6, find the probability that Bob loses all of his money.

Solution. Let one unit be \$5. Then Alice starts with $i = 5$ units and Bob starts with 10 units, so the total is $k = 15$ units.

Bob loses all of his money exactly when Alice reaches 15 units before hitting 0, so this is the gambler's ruin probability $\mathbb{P}_i$ with $i = 5$ and $k = 15$.

For (1), $p = q = 1/2$, so

$$\mathbb{P}(\text{Bob loses}) = \mathbb{P}_5 = \frac{5}{15} = \frac{1}{3}.$$

For (2), $p = 0.6$ and $q = 0.4$, so

$$\mathbb{P}(\text{Bob loses}) = \mathbb{P}_5 = \frac{1-(q/p)^5}{1-(q/p)^{15}} = \frac{1-(2/3)^5}{1-(2/3)^{15}} \approx 0.8703.$$

□

Example 22.3 (Reflecting random walks) This is a regular random walk on the non-negative integers with a reflecting barrier at 0. If $X_n = 0$, then $X_{n+1} = 1$. To analyze λ_{00} = "return to 0", note that $T_{00} = T_{01} + T_{10}$ since we must first go to 1 and then return to 0. Because we always move from 0 to 1 in one trial, $T_{00} = 1 + T_{10}$. Also, λ_{10} in this random walk is exactly the same as λ_{10} in the original random walk.

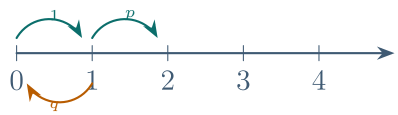

FIGURE 4

The transition structure is shown in Figure 4. Therefore T_{10} is short proper if $p < 1/2$, null proper if $p = 1/2$, and improper if $p > 1/2$. So λ_{00} is positive recurrent if $p < 1/2$, null recurrent if $p = 1/2$, and transient if $p > 1/2$.

Example 22.4 (Random walk on a circle of size m.) If $m = 2$, the process in Figure 5 is

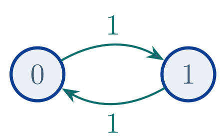

FIGURE 5

which is deterministic alternation between 0 and 1. For $m = 6$, the return probabilities are more complicated. For λ_{00},

$$r_6 = \binom{6}{3} p^3 q^3 + p^6 + q^6.$$

To analyze this further, we will need Markov chains.

Lecture 23 — Wednesday, July 02, 2014

7. Discrete-Time Markov Chains

Introduction to Markov Chains

Definition 23.1 A **discrete-time Markov chain** is a stochastic process $\{X_n \mid n = 0, 1, 2, \dots\}$ with a countable state space S and which satisfies the **Markov property**:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i),$$

whenever the conditioning history has positive probability. In words, the future depends only on the present and *not* on the past.

We also assume **time homogeneity**, meaning $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ does not depend on n. In this case, write

$$\mathbb{P}_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i),$$

called the **one-step transition probability** from i to j. These probabilities exist for all pairs (i, j) with $i, j \in S$, and are collected into a matrix

$$\mathbf{P} = \begin{bmatrix} \mathbb{P}_{00} & \mathbb{P}_{01} & \cdots \\ \mathbb{P}_{10} & \mathbb{P}_{11} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

called the **one-step transition probability matrix**.

The matrix $\mathbf{P}$ is square ($|S| \times |S|$, finite or infinite), all entries satisfy $\mathbb{P}_{ij} \geq 0$, and each row sums to 1, which makes it a **stochastic matrix**. For example, if $S = \{1, \dots, k\}$, then

$$\sum_{j \in S} \mathbb{P}_{ij} = 1$$

for every i, since at step $n+1$ the chain must be in some state in S. In fact, any matrix satisfying these three conditions (square, nonnegative entries, row sums equal to 1) is the one-step transition matrix for some Markov chain.

Markov chains can model weather forecasts, stock prices, interest rates, bond ratings, academic progression, life status under insurance contracts, elevator motion, note progressions for a composer, and many other systems.

Example 23.2 $S = \{0, 1\}$ (which could be win/lose, rain/no rain, and so on). Then

$$\mathbf{P} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}$$

where $\mathbb{P}_{00} = \alpha \in [0, 1]$ and $\mathbb{P}_{11} = \beta \in [0, 1]$.

Example 23.3 Suppose that the wins and losses of a team depend on the outcomes of the last two games:

$n-1$	n	$\mathbb{P}$(W next)
W	W	0.8
L	W	0.5
W	L	0.3
L	L	0.1

We want to model this with a Markov chain, but a two-state model would violate the Markov property. To overcome this, define a new state space $S = \{\mathrm{WW}, \mathrm{LW}, \mathrm{WL}, \mathrm{LL}\}$. Then

$$\mathbf{P} = \begin{bmatrix} 0.8 & 0 & 0.2 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0 & 0.1 & 0 & 0.9 \end{bmatrix}.$$

Example 23.4 For simple random walk, $S = \{\ldots, -1, 0, 1, 2, \ldots\}$, so

$$\mathbf{P} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \iddots \\ \ldots & 0 & p & 0 & \ldots \\ \ldots & q & 0 & p & \ldots \\ \ldots & 0 & q & 0 & \ldots \\ \iddots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

with $\mathbb{P}_{i,i+1} = p$ and $\mathbb{P}_{i,i-1} = q$.

Lecture 24 — Wednesday, July 02, 2014

Recall that a Markov chain is a stochastic process $\{X_n \mid n = 0, 1, 2, \dots\}$ on a discrete state space S with the Markov property

$$\mathbb{P}_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i).$$

The matrix $\mathbf{P}$ consists of the $\mathbb{P}_{ij}$ values. If X_0 is known, each row of $\mathbf{P}$ is the probability mass function for the next step. We now ask about two-step and, more generally, n-step transitions.

Definition 24.1 Define

$$\mathbb{P}_{ij}^{(n)} \coloneqq \mathbb{P}(X_n = j \mid X_0 = i),$$

the ***n*-step transition probability**. Also,

$$\mathbb{P}_{ij}^{(n)} = \mathbb{P}(X_{n+m} = j \mid X_m = i).$$

The matrix $\mathbf{P}^{(n)}$ contains all values $\mathbb{P}_{ij}^{(n)}$ and has the same structural properties as $\mathbf{P}$. For $n = 1$, $\mathbf{P}^{(1)} = \mathbf{P}$, and for $n = 0$, $\mathbf{P}^{(0)} = I$.

Theorem 24.2 (Chapman–Kolmogorov equations) These give a way to obtain n-step transition probabilities from one-step transition probabilities:

$$\mathbb{P}_{ij}^{(n+m)} = \sum_{k \in S} \mathbb{P}_{ik}^{(n)} \cdot \mathbb{P}_{kj}^{(m)}.$$

There are many forms of the Chapman–Kolmogorov equations. This is the pointwise form.

Proof.

$$\begin{aligned}
\mathbb{P}_{ij}^{(n+m)} &= \mathbb{P}(X_{n+m} = j \mid X_0 = i) = \sum_{k \in S} \mathbb{P}(X_{n+m} = j, X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_n = k \mid X_0 = i) \cdot \mathbb{P}(X_{n+m} = j \mid X_n = k, X_0 = i) \quad \text{(by Markov property)} \\
&= \sum_{k \in S} \mathbb{P}_{ik}^{(n)} \cdot \mathbb{P}(X_{n+m} = j \mid X_n = k) = \sum_{k \in S} \mathbb{P}_{ik}^{(n)} \cdot \mathbb{P}_{kj}^{(m)}.
\end{aligned}$$

□

In matrix form, the Chapman–Kolmogorov equations become

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)}\mathbf{P}^{(m)}.$$

In particular,

$$\mathbf{P}^{(1)} = \mathbf{P}, \qquad \mathbf{P}^{(2)} = \mathbf{P}\mathbf{P} = \mathbf{P}^2, \qquad \mathbf{P}^{(3)} = \mathbf{P}^{(2)}\mathbf{P}^{(1)} = \mathbf{P}^3, \qquad \ldots, \qquad \text{and} \qquad \mathbf{P}^{(n)} = \mathbf{P}^n.$$

Thus $\mathbf{P}$ contains all information about the chain's future movement: if $X_0 = i$, then the distribution of X_n is the ith row of $\mathbf{P}^n$.

Renewal Theory Applied to Markov Chains

Any discrete Markov chain is a discrete stochastic process. For states $i, j \in S$, define $\lambda_{ii} :=$ "return to state i" and $\lambda_{ij} :=$ "visit state j, starting from i". The event λ_{ii} is a renewal event (by the Markov property and time homogeneity), and λ_{ij} is a delayed renewal event with associated renewal event λ_{jj}. Hence the event classifications from renewal theory for λ_{ii} apply to state i in the chain.

Definition 24.3 Let T_{ii} be the number of steps to first return to i. If $f_{ii} := \mathbb{P}(T_{ii} < \infty) = 1$, then state i is **recurrent**. If $f_{ii} < 1$, then i is **transient**. A recurrent state is **positive recurrent** if $\mathbb{E}[T_{ii}] < \infty$ and **null recurrent** if $\mathbb{E}[T_{ii}] = \infty$.

The **period** of state i is

$$d(i) = \gcd\{n \geq 1 : \mathbb{P}_{ii}^{(n)} > 0\}.$$

The state is **aperiodic** if $d(i) = 1$ and **periodic** otherwise.

Each state has its own renewal sequence:

$$r_n = \mathbb{P}(\lambda_{ii} \text{ occurs on trial } n) = \mathbb{P}(X_n = i \mid X_0 = i) = \mathbb{P}_{ii}^{(n)},$$

which is the ith diagonal element of $\mathbf{P}^{(n)} = \mathbf{P}^n$.

If $\lim_{n\to\infty} \mathbb{P}_{ii}^{(n)} = p \neq 0$, then by the aperiodic renewal theorem,

$$\mathbb{E}[T_{ii}] = \frac{1}{p}.$$

If the event is periodic with period d, then

$$\mathbb{E}\left[T_{ii}\right] = \frac{d}{\lim_{n\to\infty} \mathbb{P}_{ii}^{(dn)}}.$$

Example 24.4

1. Consider random walk with $S = \mathbb{Z}$. If $p = 1/2$, state 0 is null recurrent and periodic with period 2. All states are null recurrent with period 2. If $p \neq 1/2$, all states are transient.
2. Consider the gambler's ruin setup. Here $\mathbb{P}_{00} = \mathbb{P}_{kk} = 1$, so $\mathbb{E}\left[T_{00}\right] = \mathbb{E}\left[T_{kk}\right] = 1$, and states 0 and k are positive recurrent. States $1, 2, \ldots, k-1$ are transient.

Lecture 25 — Thursday, July 03, 2014

Communication and Classes in Markov Chains

Recall that we say that state i in a Markov chain is recurrent, transient, and so on according to the event λ_{ii} = "return to i". Some chains have all states of the same type, while others do not; this is explained by class structure.

Definition 25.1 State j is **accessible** from state i (written $i \to j$) if there is a path from i to j (that is, if there exists n such that $\mathbb{P}_{ij}^{(n)} > 0$).

Note that $\mathbb{P}_{ij} > 0$ is sufficient but not necessary for accessibility.

Definition 25.2 States i and j **communicate** (written $i \leftrightarrow j$) if i is accessible from j and j is accessible from i (that is, if there exist m, n such that $\mathbb{P}_{ij}^{(n)} > 0$ and $\mathbb{P}_{ji}^{(m)} > 0$).

Communication partitions the state space S into disjoint **classes**: states in the same class communicate, and states in different classes do not.

Example 25.3 For

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix},$$

there are two classes: $C_1 = \{0\}$ and $C_2 = \{1, 2\}$.

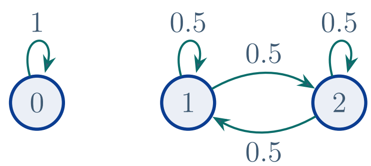

FIGURE 6

Figure 6 displays the two closed classes. By contrast, Figure 7 shows a three-state directed cycle:

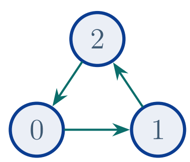

FIGURE 7

Then all states communicate and there is only one class.

There are two broad types of classes in Markov chains:

1. **Closed classes:** $\mathbb{P}_{ij} = 0$ for all $i \in C$ and $j \notin C$ (there is no way to leave the class).
2. **Open classes:** there exists $i \in C$ and $j \notin C$ with $\mathbb{P}_{ij} > 0$ (we can leave the class, but not come back). If we could come back, then j and i would communicate, so $j \in C$, a contradiction.

Every finite Markov chain has at least one closed class. An infinite chain need not: for example, the deterministic chain on $\mathbb{N}$ that moves from i to $i + 1$ has none.

Definition 25.4 If a Markov chain has only one class (equivalently, all states communicate), then the chain is **irreducible**.

Example 25.5

1. In simple random walk, for any (i, j), we can move from i to j and from j to i, so $i \leftrightarrow j$. Hence all states lie in one class, and the random walk is irreducible.
2. In the gambler's ruin setup,

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ q & 0 & p & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Here $C_1 = \{0\}$ is closed, $C_2 = \{k\}$ is closed, and $C_3 = \{1, 2, \dots, k-1\}$ is open.

Theorem 25.6 All states in the same class of a Markov chain are of the same type (transient, null recurrent, or positive recurrent) and have the same period.

Proof. Let i and j be in the same class. Then $i \leftrightarrow j$, so there exist integers $m, n \geq 1$ such that $\mathbb{P}_{ij}^{(m)} > 0$ and $\mathbb{P}_{ji}^{(n)} > 0$.

For every $r \geq 0$, we can go from j to i in n steps, return from i to i in r steps, and then go from i to j in m steps. Therefore,

$$\mathbb{P}_{jj}^{(m+n+r)} \geq \mathbb{P}_{ji}^{(n)} \mathbb{P}_{ii}^{(r)} \mathbb{P}_{ij}^{(m)}.$$

Summing over r gives

$$\sum_{t \geq 0} \mathbb{P}_{jj}^{(t)} \geq \mathbb{P}_{ji}^{(n)} \mathbb{P}_{ij}^{(m)} \sum_{r \geq 0} \mathbb{P}_{ii}^{(r)}.$$

By symmetry, the same argument with i and j interchanged gives

$$\sum_{t \geq 0} \mathbb{P}_{ii}^{(t)} \geq \mathbb{P}_{ij}^{(m)} \mathbb{P}_{ji}^{(n)} \sum_{r \geq 0} \mathbb{P}_{jj}^{(r)}.$$

Hence one of these series is finite if and only if the other is finite. So i is transient if and only if j is transient, and i is recurrent if and only if j is recurrent.

Now suppose i and j are recurrent, and write $\mu_i = \mathbb{E}\left[T_{ii}\right]$ and $\mu_j = \mathbb{E}\left[T_{jj}\right]$. Assume $\mu_i < \infty$. Let τ_{ij} be the first hitting time of j starting from i. Since $i \to j$, there is a path from i to j, so

$$a \coloneqq \mathbb{P}_i(\tau_{ij} < T_{ii}) > 0.$$

Each excursion from i back to i hits j before returning with probability at least a, so the number of excursions needed to hit j has finite mean. Because each excursion has mean length μ_i, we get $\mathbb{E}_i[\tau_{ij}] < \infty$. Moreover, on a fixed positive-probability path from i to j before returning to i, the remaining time to return to i has mean $\mathbb{E}_j[\tau_{ji}]$. Since $\mu_i < \infty$, this forces $\mathbb{E}_j[\tau_{ji}] < \infty$.

Starting from j, a return to j occurs no later than going from j to i and then from i back to j. Using the restart principle described above,

$$\mu_j = \mathbb{E}_j\left[T_{jj}\right] \leq \mathbb{E}_j\left[\tau_{ji}\right] + \mathbb{E}_i\left[\tau_{ij}\right] < \infty.$$

Thus j is positive recurrent. Interchanging i and j shows the converse, so recurrent states in one class are either all positive recurrent or all null recurrent.

Finally, let $D_i = \{r \geq 1 : \mathbb{P}_{ii}^{(r)} > 0\}$ and $D_j = \{r \geq 1 : \mathbb{P}_{jj}^{(r)} > 0\}$. Set $c = m + n$. The paths between i and j show that $c \in D_i \cap D_j$, each $r \in D_i$ gives $c + r \in D_j$, and each $s \in D_j$ gives $c + s \in D_i$. Hence $\gcd(D_j)$ divides both $c + r$ and c, so it divides every $r \in D_i$; therefore $\gcd(D_j) \mid \gcd(D_i)$. The symmetric argument gives the reverse divisibility. Thus

$$d(i) = \gcd(D_i) = \gcd(D_j) = d(j).$$

Hence all states in the same class have the same type and the same period. $\square$

Example 25.7 True or false: if $\mathbb{P}_{ij} = 0 = \mathbb{P}_{ji}$, then $i \not\leftrightarrow j$.

Solution. False! $\square$

Lecture 26 — Monday, July 07, 2014

Theorem 26.1 Every state in an open class is transient.

Proof. Because C is open, some finite path leaves C with positive probability. Fix $k \in C$. Since every pair of states in C communicates, we can choose a finite path from k to the start of that exit path which does not revisit k after leaving it. Following the combined path exits C before returning to k, and this event has positive probability. Once the chain leaves C, it cannot return to k. Hence

$$\mathbb{P}_k(T_{kk} = \infty) > 0,$$

so $f_{kk} < 1$ and k is transient. □

Theorem 26.2 Finite closed classes contain positive recurrent states.

Proof. Fix $i \in C$. Because C is finite and irreducible, from every $k \in C$ there is a positive-probability path to i; when $k = i$, choose a positive-length return cycle. We can therefore choose $L < \infty$ and $\varepsilon > 0$ so that, from any state in C, the chain hits i after a positive number of steps but within the next L steps with probability at least ε. Since C is closed, the chain cannot escape while it waits. The Markov property now gives

$$\mathbb{P}_i(T_{ii} > mL) \leq (1 - \varepsilon)^m.$$

This geometric tail implies $\mathbb{E}_i[T_{ii}] < \infty$. Thus i, and hence every state in C, is positive recurrent. □

Theorem 26.3 An infinite closed class may be transient, null recurrent, or positive recurrent.

Proof. For a simple random walk, $p = 1/2$ implies all states are null recurrent, while $p \neq 1/2$ implies all states are transient. For a random walk with a reflecting barrier at 0, all states still communicate and the state space is infinite. If $p < 1/2$, all states are positive recurrent; if $p = 1/2$, they are null recurrent; and if $p > 1/2$, they are transient. □

Periodicity is also a class property: if i has period d and $j \leftrightarrow i$, then j also has period d. We can determine a class period from the state diagram. If any $\mathbb{P}_{ii} > 0$, then $d = 1$. If all $\mathbb{P}_{ii} = 0$, the class may or may not be periodic.

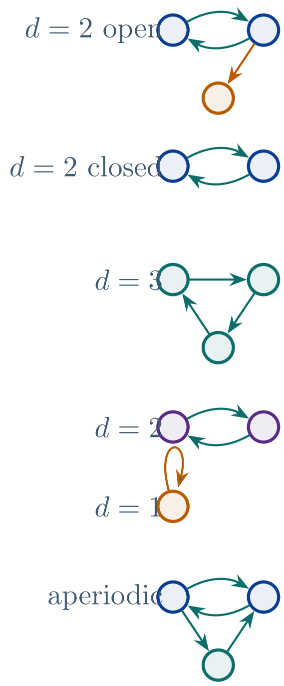

FIGURE 8

Figure 8 collects representative class periods.

Markov Chains in the Long Run

If we know X_0, we can find the probability of being in any state at time n. We may also want a formulation that starts from a distribution on X_0.

Definition 26.4 Let $\boldsymbol{p}_n = (p_{n0}, p_{n1}, \ldots, p_{nj}, \ldots)$ be a $1 \times |S|$ row vector containing the probability mass function of X_n, where $p_{ni} = \mathbb{P}(X_n = i)$. Its entries are nonnegative and sum to 1.

By the Chapman–Kolmogorov equations,

$$\mathbb{P}(X_{n+1} = j) = p_{n+1,j} = \sum_{\text{all } k} p_{nk} \cdot \mathbb{P}_{kj} = \sum_{\text{all } k} \mathbb{P}(X_n = k) \cdot \mathbb{P}(X_{n+1} = j \mid X_n = k).$$

In matrix form,

$$\boldsymbol{p}_{n+1} = \boldsymbol{p}_n \mathbf{P}.$$

Definition 26.5 A probability vector $\boldsymbol{\pi} = (\pi_0, \pi_1, \dots)$ is a **stationary distribution** if

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}.$$

Thus, if $\boldsymbol{p}_0 = \boldsymbol{\pi}$, then $\boldsymbol{p}_n = \boldsymbol{\pi}$ for every n.

A limiting distribution is a separate notion: in favourable cases, $\boldsymbol{p}_n \to \boldsymbol{\pi}$ from any initial distribution. For an irreducible positive recurrent chain, π_j is the long-run proportion of time spent in state j, and

$$\mathbb{E}_j\left[T_{jj}\right] = \frac{1}{\pi_j}.$$

The matrix powers converge only under the additional aperiodicity condition developed below.

Lecture 27 — Tuesday, July 08, 2014

Stationary Distributions and Doubly Stochastic Chains

Recall that if a vector π satisfies (1) all entries are nonnegative, (2) entries sum to 1, and (3) $\pi\mathbf{P} = \pi$, then π is a stationary/equilibrium distribution.

When does $\boldsymbol{\pi}$ exist uniquely?

1. If all states communicate (i.e., the chain is irreducible) and are positive recurrent, then $\boldsymbol{\pi}$ is guaranteed to exist and be unique. To find it, we can guess and verify, or solve $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ with

$$\sum_{i \in S} \pi_i = 1.$$

Example 27.1 Let $0 \leq \alpha, \beta < 1$, and let

$$\mathbf{P} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}.$$

Solving $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ gives

$$\pi_0 = \frac{1-\beta}{(1-\alpha)+(1-\beta)} \qquad \text{and} \qquad \pi_1 = \frac{1-\alpha}{(1-\alpha)+(1-\beta)}.$$

If a finite n-state transition matrix has columns that sum to 1, then $\mathbf{P}$ is **doubly stochastic** and $\boldsymbol{\pi}$ is uniform:

$$\boldsymbol{\pi} = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Example 27.2 Consider random walk on a circle with

$$\mathbf{P} = \begin{bmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{bmatrix}.$$

and $q = 1-p$. Then $\mathbf{P}$ is doubly stochastic, so $\boldsymbol{\pi}$ is uniform.

2. In any stationary probability distribution, every transient or null recurrent state i has $\pi_i = 0$.
3. If a Markov chain has multiple positive recurrent closed classes, each such class has its own stationary distribution. Restrict to class C and solve

$$\boldsymbol{\pi}_C \mathbf{P}_C = \boldsymbol{\pi}_C,$$

where $\mathbf{P}_C$ is the transition matrix within C and $\boldsymbol{\pi}_C$ contains the components indexed by states in C.

If, in addition to positive recurrence, all states in an irreducible chain are aperiodic, then

$$\lim_{n\to\infty} \mathbf{P}^n = \boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{bmatrix} = \begin{bmatrix} \pi_0 & \cdots & \pi_k \\ \vdots & \ddots & \vdots \\ \pi_0 & \cdots & \pi_k \end{bmatrix},$$

so the limiting distribution is the same regardless of starting state.

Hence π_i is the long-run proportion of time spent in state i. If the chain is periodic, the matrix powers need not converge, although the unique stationary distribution still exists. The Cesàro averages do converge:

$$\frac{1}{n}\sum_{m=0}^{n-1}\mathbf{P}^m \longrightarrow \mathbf{\Pi},$$

and the empirical time proportions converge to $\boldsymbol{\pi}$. In either case, $\mathbb{E}_i[T_{ii}] = 1/\pi_i$. If there is more than one closed class, the starting state matters.

Example 27.3 Consider

$$S = \{0, 1, \ldots, 6\} \qquad \text{and} \qquad \mathbf{P} = \begin{bmatrix} 1/3 & 0 & 0 & 0 & 2/3 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 3/4 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/8 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 1/4 & 0 & 0 & 1/2 \end{bmatrix}.$$

Find the classes of the chain, and determine the associated stationary distributions.

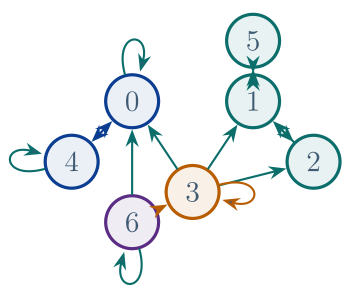

FIGURE 9

Solution. Figure 9 shows the class structure of the chain. The classes are

$$C_1 = \{0, 4\} \text{ closed}, \qquad C_2 = \{1, 2, 5\} \text{ closed}, \qquad C_3 = \{3\} \text{ open}, \qquad \text{and} \qquad C_4 = \{6\} \text{ open}.$$

We seek $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots, \pi_6) = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)$, with $\pi_3 = 0$ and $\pi_6 = 0$.

For C_1, we solve

$$(\pi_0, \pi_4)\begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix} = (\pi_0, \pi_4).$$

The columns sum to 1, so this is doubly stochastic and $\pi_0 = \pi_4 = 1/2$, so

$$\boldsymbol{\pi}_{C_1} = \left(\frac{1}{2}, \frac{1}{2}\right).$$

For C_2:

$$\mathbf{P}_{C_2} = \begin{bmatrix} 0 & 1/4 & 3/4 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

This is not doubly stochastic, and it has period 2. Logically, $\pi_1 = 1/2$ (every other step is in state 1), and $\pi_5 = 3\pi_2$. Hence

$$\boldsymbol{\pi}_{C_2} = \left(\frac{1}{2}, \frac{1}{8}, \frac{3}{8}\right).$$

□

Lecture 28 — Thursday, July 10, 2014

In Example 27.3, we found that $C_1 = \{0, 4\}$ with $(\pi_0, \pi_4) = (1/2, 1/2)$, and $C_2 = \{1, 2, 5\}$ with $(\pi_1, \pi_2, \pi_5) = (1/2, 1/8, 3/8)$. For transient states 3 and 6, $\pi_3 = 0$ and $\pi_6 = 0$. How do we combine these pieces of information?

We cannot simply set

$$\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6) = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{8}, 0, \frac{1}{2}, \frac{3}{8}, 0\right),$$

because these entries do not sum to 1. The key point is that the starting state X_0 matters when there are multiple closed classes.

If $X_0 \in \{0, 4\}$, then

$$\boldsymbol{\pi}_{(1)} = \left(\frac{1}{2}, 0, 0, 0, \frac{1}{2}, 0, 0\right).$$

If $X_0 \in \{1, 2, 5\}$, then

$$\boldsymbol{\pi}_{(2)} = \left(0, \frac{1}{2}, \frac{1}{8}, 0, 0, \frac{3}{8}, 0\right).$$

Both satisfy $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, so there are at least two distinct stationary solutions. In fact, for any

$\alpha \in [0,1]$, the convex combination

$$\alpha \boldsymbol{\pi}_{(1)} + (1-\alpha)\boldsymbol{\pi}_{(2)}$$

is also a solution. Thus there are infinitely many solutions.

What if $X_0 = 3$ or 6? Then we compute absorption probabilities from transient states into each closed class C_j.

Definition 28.1 The **absorption probability** A_{k,C_j} is the probability that the Markov chain ends up in the closed class C_j, given that $X_0 = k$, where k is a transient state.

In Example 27.3, we compute A_{6,C_2}.

$$\begin{aligned}
A_{6,C_2} &= \mathbb{P}(\text{start in } 6,\ \text{end up in class } C_2 = \{1,2,5\}) \\
&= \mathbb{P}(\text{go straight there}) + \mathbb{P}(\text{go to transient state first, and then } C_2) \\
&= \sum_{i \in C_2} \mathbb{P}_{6,i} + \sum_{\text{all transient } k} \mathbb{P}_{6,k} A_{k,C_2} \\
&= \mathbb{P}_{6,1} + \mathbb{P}_{6,2} + \mathbb{P}_{6,5} + \mathbb{P}_{6,3} A_{3,C_2} + \mathbb{P}_{6,6} A_{6,C_2} \\
&= 0 + 0 + 0 + \frac{1}{4} A_{3,C_2} + \frac{1}{2} A_{6,C_2}
\end{aligned}$$

We need a second equation:

$$\begin{aligned}
A_{3,C_2} &= \mathbb{P}_{3,1} + \mathbb{P}_{3,2} + \mathbb{P}_{3,5} + \mathbb{P}_{3,3} A_{3,C_2} + \mathbb{P}_{3,6} A_{6,C_2} \\
&= \frac{1}{4} + \frac{1}{4} + 0 + \frac{1}{4} A_{3,C_2} + 0
\end{aligned}$$

Solving, we get $A_{3,C_2} = 2/3$ and $A_{6,C_2} = 1/3$. Note that these do not have to add up to 1, since they have different conditions. What we do need is for $A_{3,C_2} + A_{3,C_1} = 1$ and $A_{6,C_2} + A_{6,C_1} = 1$, since the chain will get absorbed into exactly one of C_1 or C_2. So $A_{3,C_1} = 1/3$ and $A_{6,C_1} = 2/3$.

These absorption-weighted mixtures describe the long-run occupation distribution. If $X_0 = 3$, it is $\frac{1}{3}\boldsymbol{\pi}_{(1)} + \frac{2}{3}\boldsymbol{\pi}_{(2)}$; if $X_0 = 6$, it is $\frac{2}{3}\boldsymbol{\pi}_{(1)} + \frac{1}{3}\boldsymbol{\pi}_{(2)}$. Because C_2 is periodic, the fixed-time distribution need not converge to these mixtures.

8. The Exponential Distribution and Poisson Processes

The Alarm Clock Lemma

The exponential distribution is the foundation for continuous-time Markov chains. The Poisson process is a pure-birth process. Let us recall some basic facts about this distribution.

We say $T \sim \text{Exp}(\lambda)$ if

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

The parameter λ is the rate (i.e., frequency of events). Also,

$$\mathbb{E}[T] = \frac{1}{\lambda} = \theta \qquad \text{and} \qquad \text{Var}(T) = \frac{1}{\lambda^2} = \theta^2,$$

so an equivalent parameterization is

$$f(t) = \frac{1}{\theta} e^{-t/\theta}.$$

Furthermore,

$$F(t) = \mathbb{P}(T \leq t) = 1 - e^{-\lambda t} \qquad \text{and} \qquad \mathbb{P}(T > t) = e^{-\lambda t}.$$

Recall the memoryless property:

$$\mathbb{P}(T > t + s \mid T > t) = \mathbb{P}(T > s).$$

Lecture 29 — Monday, July 14, 2014

Recall from last class that for $T \sim \text{Exp}(\lambda)$,

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0, \qquad F(t) = 1 - e^{-\lambda t}, \qquad \mathbb{E}[T] = \frac{1}{\lambda}, \qquad \text{and} \qquad \text{Var}(T) = \frac{1}{\lambda^2}.$$

It is also easy to show that the mgf of T is

$$M(s) = \frac{\lambda}{\lambda - s}, \qquad s < \lambda.$$

Theorem 29.1 (Alarm clock lemma) Let $T_1, \ldots, T_n$ be independent random variables with rates λ_i, and $T_i \sim \text{Exp}(\lambda_i)$. This is the picture of n independent alarm clocks, each set to ring at a random future time.

1. Let $T = \min\{T_1, \ldots, T_n\}$. Then

$$T \sim \text{Exp}\left(\sum_{i=1}^{n} \lambda_i\right)$$

is the time you wake up.

2.

$$\mathbb{P}(T_j = T) = \frac{\lambda_j}{\sum_{i=1}^{n} \lambda_i}.$$

3. The event "$T_j = T$" is independent of the actual value of T.

Proof. For (1),

$$\begin{aligned}
\mathbb{P}(T \le t) &= 1 - \mathbb{P}(T > t) \\
&= 1 - \mathbb{P}(\min\{T_1, \ldots, T_n\} > t) \\
&= 1 - \mathbb{P}(T_1 > t, \ldots, T_n > t) \\
&\overset{\text{independence}}{=} 1 - \prod_{i=1}^{n} \mathbb{P}(T_i > t) \\
&= 1 - e^{-\lambda_1 t} \cdots e^{-\lambda_n t} \\
&= 1 - e^{-t \sum_i \lambda_i}.
\end{aligned}$$

This is the cdf of the

$$\text{Exp}\left(\sum_{i=1}^{n} \lambda_i\right)$$

distribution.

For (2), we show

$$\mathbb{P}(T_j = T) = \frac{\lambda_j}{\sum_{i=1}^{n} \lambda_i}.$$

For fixed j,

$$\mathbb{P}(T_j = T) = \mathbb{P}(T_j \le T_i \ \forall i \ne j) = \mathbb{P}(T_i \ge T_j \ \forall i \ne j).$$

$$\begin{aligned}
\mathbb{P}(T_j = T) &= \int_0^\infty \mathbb{P}(T_i \ge t \ \forall i \ne j \mid T_j = t)\, f_{T_j}(t)\, \mathrm{d}t \\
&= \int_0^\infty \mathbb{P}(T_i \ge t \ \forall i \ne j)\, f_{T_j}(t)\, \mathrm{d}t
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \prod_{i \neq j} \mathbb{P}(T_i \geq t)\, \lambda_j e^{-\lambda_j t}\, \mathrm{d}t \\
&= \int_0^\infty \lambda_j e^{-t \sum_{i=1}^n \lambda_i}\, \mathrm{d}t \\
&= \frac{\lambda_j}{\sum_{i=1}^n \lambda_i}.
\end{aligned}
$$

For (3), let $A_j = \{T_j = T\}$. Then

$$
\mathbb{P}(A_j, T \in \mathrm{d}t) = \lambda_j e^{-t \sum_{i=1}^n \lambda_i}\, \mathrm{d}t \qquad \text{and} \qquad f_T(t) = \left(\sum_{i=1}^n \lambda_i \right) e^{-t \sum_{i=1}^n \lambda_i}.
$$

Therefore

$$
\mathbb{P}(A_j \mid T = t) = \frac{\lambda_j e^{-t \sum_{i=1}^n \lambda_i}}{\left(\sum_{i=1}^n \lambda_i\right) e^{-t \sum_{i=1}^n \lambda_i}} = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i},
$$

which is independent of t. So A_j is independent of T. □

There are other ways to combine expressions.

1. The convolution of independent and identically distributed exponential random variables: if $T_1, \ldots, T_n$ are independent and identically distributed with $T_i \sim \mathrm{Exp}(\lambda)$, then

$$
T = T_1 + \cdots + T_n \sim \mathrm{Gamma}(n, \lambda),
$$

   with

$$
f_T(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{\Gamma(n)} \qquad \text{and} \qquad \Gamma(n) = (n-1)!, \quad n \in \mathbb{Z}^+,
$$

   and moment generating function

$$
M_T(s) = \left(\frac{\lambda}{\lambda - s} \right)^n.
$$

Proof.

$$\begin{aligned}
M_T(s) &= \mathbb{E}\left[e^{Ts}\right] \\
&= \mathbb{E}\left[e^{(T_1+\cdots+T_n)s}\right] \\
&\overset{\text{independence}}{=} \mathbb{E}\left[e^{T_1 s}\right]\cdots\mathbb{E}\left[e^{T_n s}\right] \\
&= M_{T_1}(s)\cdots M_{T_n}(s) \\
&= \left(\frac{\lambda}{\lambda - s}\right)^n .
\end{aligned}$$

This is the mgf of the Gamma(n, λ) distribution. □

2. If $T_1, \ldots, T_n$ are exponentially distributed but with different rates, then the distribution of $T = T_1 + \cdots + T_n$ has a **hypoexponential** distribution (which is not covered in this course).

Counting Processes and Poisson Processes

Definition 29.2 A **counting process** is a continuous-time stochastic process $\{N(t), t \geq 0\}$ which satisfies:

1. $N(t) \geq 0$,
2. $N(t) \in \mathbb{Z}$ for all t,
3. $N(t)$ is nondecreasing in t, and
4. $N(t) - N(s)$ is the number of events in the interval $(s, t]$.

Lecture 30 — Wednesday, July 16, 2014

Recall that a counting process $\{N(t), t \geq 0\}$ satisfies $N(t) \in \mathbb{Z}_{\geq 0}$ and is nondecreasing; the number of events in $(s, t]$ is $N(t) - N(s)$.

Definition 30.1 $N(t+s) - N(t)$ counts events in $(t, t+s]$, where s is the interval width. The process has **stationary increments** if the distribution of this increment depends only on s, not on t. It has **independent increments** if the increments over every finite collection of pairwise disjoint time intervals are mutually independent.

Poisson Processes

Definition 30.2 A counting process $\{N(t), t \geq 0\}$ is a **Poisson process** if:

1. $N(0) = 0$,
2. It has stationary and independent increments, and
3. Its interarrival times are independent and identically distributed as $\text{Exp}(\lambda)$ for some common rate $\lambda > 0$.

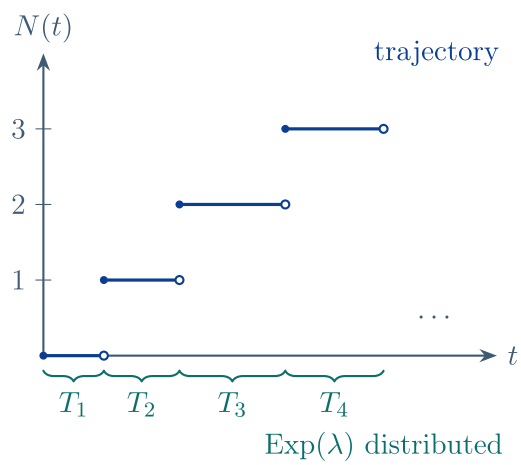

FIGURE 10

Figure 10 shows a typical counting path and its interarrival times.

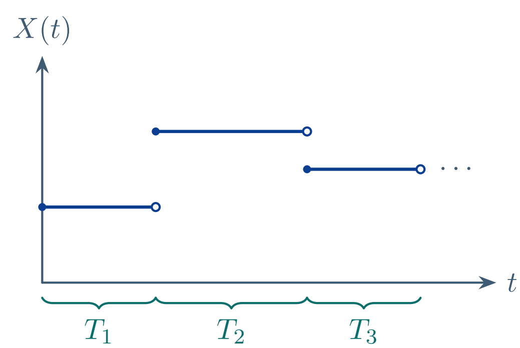

FIGURE 11

Figure 11 shows a more general right-continuous jump path. There are several equivalent definitions of the Poisson process.

Theorem 30.3 Fix $\lambda > 0$. Let $\{N(t), t \geq 0\}$ be a counting process with $N(0) = 0$, stationary increments, and independent increments. The following are equivalent:

1. The interarrival times $T_1, T_2, \ldots$ are independent and identically distributed as $\text{Exp}(\lambda)$.
2. For $h \downarrow 0$,
$$\mathbb{P}(N(t+h) - N(t) = 1) = \lambda h + o(h) \qquad \text{and} \qquad \mathbb{P}(N(t+h) - N(t) \geq 2) = o(h).$$
3. For every $t \geq 0$, $N(t) \sim \text{Poisson}(\lambda t)$.
4. For every $k \in \mathbb{Z}^+$, the waiting time to the k^{th} event,
$$S_k := \inf\{t \geq 0 : N(t) = k\},$$
has a $\text{Gamma}(k, \lambda)$ distribution.

Proof. We prove a chain of implications that yields equivalence of all four statements.

(1) $\Rightarrow$ (2)**.** Let $\Delta_h N(t) := N(t+h) - N(t)$. By stationary increments, it is enough to work at time 0. Since $T_1 \sim \text{Exp}(\lambda)$,

$$\mathbb{P}(\Delta_h N(0) = 0) = \mathbb{P}(T_1 > h) = e^{-\lambda h} = 1 - \lambda h + o(h),$$

so

$$\mathbb{P}(\Delta_h N(0) \geq 1) = \lambda h + o(h).$$

Also,

$$\mathbb{P}(\Delta_h N(0) \geq 2) = \mathbb{P}(S_2 \leq h),$$

and $S_2 \sim \text{Gamma}(2, \lambda)$, hence

$$\mathbb{P}(S_2 \leq h) = \int_0^h \lambda^2 s e^{-\lambda s} \, \mathrm{d}s = O(h^2) = o(h).$$

Therefore,

$$\mathbb{P}(\Delta_h N(0) = 1) = \mathbb{P}(\Delta_h N(0) \geq 1) - \mathbb{P}(\Delta_h N(0) \geq 2) = \lambda h + o(h).$$

By stationarity, the same formulas hold for $\Delta_h N(t)$ for every $t \geq 0$.

(2) $\Rightarrow$ (3). Define $p_n(t) := \mathbb{P}(N(t) = n)$ for $n \in \mathbb{Z}_{\geq 0}$. Using independent increments,

$$p_n(t+h) = \sum_{j=0}^{n} p_{n-j}(t)\,\mathbb{P}(\Delta_h N(t) = j),$$

where $\Delta_h N(t) := N(t+h) - N(t)$. By the infinitesimal assumptions,

$$\mathbb{P}(\Delta_h N(t) = 0) = 1 - \lambda h + o(h),$$
$$\mathbb{P}(\Delta_h N(t) = 1) = \lambda h + o(h) \qquad \text{and} \qquad \mathbb{P}(\Delta_h N(t) \geq 2) = o(h).$$

Thus, for $n \geq 1$,

$$p_n(t+h) - p_n(t) = \lambda h\, p_{n-1}(t) - \lambda h\, p_n(t) + o(h),$$

and for $n = 0$,

$$p_0(t+h) - p_0(t) = -\lambda h\, p_0(t) + o(h).$$

Dividing by h and sending $h \downarrow 0$ gives the **Kolmogorov forward equations**

$$p_0'(t) = -\lambda p_0(t) \qquad \text{and} \qquad p_0(0) = 1,$$

and

$$p_n'(t) = \lambda p_{n-1}(t) - \lambda p_n(t) \qquad \text{and} \qquad p_n(0) = 0 \text{ for } n \geq 1.$$

Solving recursively yields

$$p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots,$$

so $N(t) \sim \text{Poisson}(\lambda t)$.

(3) $\Rightarrow$ (4). For $k \in \mathbb{Z}^+$,

$$\{S_k \leq t\} = \{N(t) \geq k\},$$

so

$$F_{S_k}(t) = \mathbb{P}(S_k \leq t) = 1 - \sum_{j=0}^{k-1} e^{-\lambda t}\frac{(\lambda t)^j}{j!}.$$

Differentiating,

$$f_{S_k}(t) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!}, \quad t > 0,$$

which is the density of the Gamma(k, λ) distribution.

(4) $\Rightarrow$ (3). The events $\{S_k \leq t\}$ and $\{S_{k+1} \leq t\}$ are nested, so for $k \geq 1$,

$$\mathbb{P}(N(t) = k) = \mathbb{P}(S_k \leq t) - \mathbb{P}(S_{k+1} \leq t).$$

Substituting the two gamma cdfs and cancelling their common terms gives

$$\mathbb{P}(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

Also,

$$\mathbb{P}(N(t) = 0) = \mathbb{P}(S_1 > t) = e^{-\lambda t}.$$

Hence, for all $k \in \mathbb{Z}_{\geq 0}$,

$$\mathbb{P}(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!},$$

so $N(t) \sim \text{Poisson}(\lambda t)$.

(3) ⇒ (1). Let $T_1 := S_1$. Then

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t},$$

so $T_1 \sim \text{Exp}(\lambda)$. For $n \geq 1$, the next waiting time is

$$T_{n+1} = S_{n+1} - S_n.$$

To pass from increments over fixed intervals to a restart at a random arrival time requires the strong Markov property, whose proof is beyond the scope of this course. The resulting restart principle says that the process after each arrival is independent of its past and has the original law. It follows that each T_{n+1} has the same $\text{Exp}(\lambda)$ law as T_1 and is independent of $(T_1, \ldots, T_n)$. Hence $T_1, T_2, \ldots$ are independent and identically distributed as $\text{Exp}(\lambda)$.

We have proved

$$(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (3),$$

and also (3) ⇒ (1). Therefore all four statements are equivalent. □

The Poisson process also has a **thinning property**. Classify events independently as type 1 or type 2, with $\mathbb{P}(\text{type } 1) = p \in [0, 1]$. Let $N_1(t)$ and $N_2(t)$ count the two types in $(0, t]$.

Proposition 30.4 (Poisson thinning) The two thinned processes are independent Poisson processes with respective rates λp and $\lambda(1 - p)$. In particular,

$$N_1(t) \sim \text{Poisson}(\lambda p t) \qquad \text{and} \qquad N_2(t) \sim \text{Poisson}(\lambda(1 - p)t).$$

Proof. For nonnegative integers a, b, condition on the total $N(t) = a + b$. Independent classification makes the type-1 count conditionally binomial, so

$$\begin{aligned}\mathbb{P}(N_1(t) = a, N_2(t) = b) &= e^{-\lambda t}\frac{(\lambda t)^{a+b}}{(a+b)!}\binom{a+b}{a}p^a(1-p)^b \\ &= \left[e^{-\lambda pt}\frac{(\lambda pt)^a}{a!}\right]\left[e^{-\lambda(1-p)t}\frac{(\lambda(1-p)t)^b}{b!}\right].\end{aligned}$$

Thus the two counts are independent and have the stated Poisson laws. Applying the same argument on every collection of disjoint time intervals, using independent increments of the original process and independent event classifications, shows that the two entire thinned processes have independent stationary increments. Hence they are independent Poisson processes. □

Lecture 31 — Thursday, July 17, 2014

Recall the Poisson process and its many interesting properties. Last time we demonstrated the thinning property.

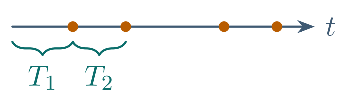

FIGURE 12

Figure 12 marks the first event times on the time axis. What is the conditional distribution of event times given the number of events that have already occurred? Given $N(t) = n$, the unordered event times are independent and uniformly distributed on $(0, t)$; equivalently, the ordered event times are order statistics from that sample.

Proposition 31.1 Let $\{N(t) : t \geq 0\}$ be a Poisson process, and let T_1 be the time that the first event occurs. For each $n \in \mathbb{Z}^+$, $t > 0$, and $0 \leq s \leq t$,

$$\mathbb{P}(T_1 \leq s \mid N(t) = n) = 1 - \left(1 - \frac{s}{t}\right)^n.$$

Equivalently,

$$\mathbb{P}(T_1 > s \mid N(t) = n) = \left(1 - \frac{s}{t}\right)^n.$$

Proof. Fix $n \in \mathbb{Z}^+$ and $0 \leq s \leq t$. The event $\{T_1 > s\}$ is the same as having no events in $(0, s]$, so

$$\mathbb{P}(T_1 > s \mid N(t) = n) = \frac{\mathbb{P}(T_1 > s,\, N(t) = n)}{\mathbb{P}(N(t) = n)} = \frac{\mathbb{P}(N(s) = 0,\, N(t) = n)}{\mathbb{P}(N(t) = n)}.$$

If $N(s) = 0$ and $N(t) = n$, then all n events occur in $(s, t]$, so

$$\mathbb{P}(N(s) = 0,\, N(t) = n) = \mathbb{P}(N(s) = 0)\, \mathbb{P}(N(t) - N(s) = n),$$

by independent increments. Using Poisson distributions for each increment,

$$\mathbb{P}(N(s) = 0) = e^{-\lambda s} \qquad \text{and} \qquad \mathbb{P}(N(t) - N(s) = n) = \frac{(\lambda(t-s))^n e^{-\lambda(t-s)}}{n!},$$

and

$$\mathbb{P}(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}.$$

Therefore,

$$\mathbb{P}(T_1 > s \mid N(t) = n) = \frac{e^{-\lambda s}\, \frac{(\lambda(t-s))^n e^{-\lambda(t-s)}}{n!}}{\frac{(\lambda t)^n e^{-\lambda t}}{n!}} = \left(\frac{t-s}{t}\right)^n = \left(1 - \frac{s}{t}\right)^n.$$

Taking complements gives

$$\mathbb{P}(T_1 \le s \mid N(t) = n) = 1 - \left(1 - \frac{s}{t}\right)^n.$$

□

Corollary 31.2 Under the same assumptions,

$$T_1 \mid N(t) = 1 \sim \text{Unif}(0, t).$$

Proof. Substitute $n = 1$ in the proposition:

$$\mathbb{P}(T_1 \le s \mid N(t) = 1) = 1 - \left(1 - \frac{s}{t}\right) = \frac{s}{t}, \quad 0 \le s \le t.$$

This is exactly the cdf of the Unif$(0, t)$ distribution. □

We can also ask about the number of events in the subinterval shown in Figure 13.

FIGURE 13

Proposition 31.3 For $t > 0$, given $N(t) = n$, the number of those n events that lie in a subinterval of length $s \le t$ is distributed as $\mathrm{Bin}(n, s/t)$.

Proof. Let $X := N(s)$, so X is the number of events in $(0, s]$. We want the conditional pmf of X given $N(t) = n$. For $k = 0, 1, \ldots, n$,

$$\mathbb{P}(X = k \mid N(t) = n) = \frac{\mathbb{P}(X = k,\, N(t) = n)}{\mathbb{P}(N(t) = n)} = \frac{\mathbb{P}(N(s) = k,\, N(t) - N(s) = n - k)}{\mathbb{P}(N(t) = n)}.$$

Because increments on disjoint intervals are independent,

$$\mathbb{P}(N(s) = k,\, N(t) - N(s) = n - k) = \mathbb{P}(N(s) = k)\mathbb{P}(N(t) - N(s) = n - k).$$

Since $N(s) \sim \mathrm{Poisson}(\lambda s)$, $N(t) - N(s) \sim \mathrm{Poisson}(\lambda(t - s))$, and $N(t) \sim \mathrm{Poisson}(\lambda t)$,

$$\begin{aligned}\mathbb{P}(X = k \mid N(t) = n) &= \frac{\dfrac{(\lambda s)^k e^{-\lambda s}}{k!} \cdot \dfrac{(\lambda(t - s))^{n-k} e^{-\lambda(t-s)}}{(n - k)!}}{\dfrac{(\lambda t)^n e^{-\lambda t}}{n!}} \\ &= \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}.\end{aligned}$$

This is exactly the pmf of the $\mathrm{Bin}(n, s/t)$ distribution. □

9. Continuous-Time Markov Chains

Many processes do not follow discrete time steps (for example, queues or times of goals scored in a game of fußbol). In a continuous-time Markov chain, the state space is still discrete, so the sample path is a right-continuous step function with jumps at random times, as in a Poisson process. Unlike the Poisson process, transitions can occur between any pair of states, not only by increments of 1.

Movement in a continuous-time Markov chain is governed by (1) waiting times between jumps, and (2) a discrete transition matrix that determines where to jump when a jump occurs. For the Markov property to hold, waiting times in (1) must be exponential with state-dependent rates. For (2), define

$$\mathbb{P}_{ij} = \mathbb{P}(\text{go to } j \mid \text{leave from } i),$$

the **embedded-chain transition probabilities**. These are nonnegative, satisfy

$$\sum_j \mathbb{P}_{ij} = 1,$$

and have $\mathbb{P}_{ii} = 0$ for every nonabsorbing state i, since a jump away from i cannot immediately land back in i. Collecting these values gives the transition matrix $\mathbf{P}$ of the embedded jump chain.

Definition 31.4 A continuous-time stochastic process $\{X(t); t \geq 0\}$ is a **continuous-time Markov chain** if:

1. The state space is countable, and
2. With $\mathcal{F}_s = \sigma(X(u) : 0 \leq u \leq s)$, it has the Markov property

$$\mathbb{P}(X(t+s) = j \mid \mathcal{F}_s) = \mathbb{P}(X(t+s) = j \mid X(s))$$

   almost surely, for every $s, t \geq 0$ and state j.

Also, we assume **time homogeneity**, so the probability

$$\mathbb{P}(X(t+s) = j \mid X(s) = i)$$

depends on t only, not on s.

Lecture 32 — Monday, July 21, 2014

Continuous-Time Generators

Recall that for a continuous-time Markov chain $\{X(t), t \geq 0\}$,

$$\mathbb{P}_{ij} = \mathbb{P}(\text{go to } j \mid \text{leaving } i) \qquad \text{and} \qquad \mathbb{P}_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i), \quad t \geq 0.$$

By the Markov property,

$$\mathbb{P}_{ij}(t) = \mathbb{P}(X(s+t) = j \mid X(s) = i).$$

If $t = 0$, then

$$\mathbb{P}_{ij}(0) = \begin{cases} 1, & j = i \\ 0, & j \neq i, \end{cases}$$

since no time has elapsed. Thus $\mathbf{P}(t) = (\mathbb{P}_{ij}(t))_{i,j\in S}$ is a family of matrices indexed by $t \geq 0$, with $\mathbf{P}(0) = \mathbb{I}_{|S|\times|S|}$. For each fixed t, $\mathbf{P}(t)$ is stochastic: its entries are nonnegative, its rows sum to 1, and the matrix is square.

The Chapman–Kolmogorov equations still hold pointwise:

$$\mathbb{P}_{ij}(t+s) = \sum_{k\in S} \mathbb{P}_{ik}(t)\mathbb{P}_{kj}(s),$$

or in matrix form,

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

Unlike the discrete time case, there is no single matrix that directly gives transition probabilities at all future times. We could choose some $\varepsilon > 0$, find $\mathbf{P}(\varepsilon)$, and then use the Chapman–Kolmogorov equations to get $\mathbf{P}(2\varepsilon), \mathbf{P}(3\varepsilon), \ldots$ but this is still discrete.

Long-run behaviour parallels the discrete time setting: we keep the same notions of transient, positive/null recurrent, accessibility, communication, open/closed classes, and irreducibility. The periodicity notion disappears because continuous-time chains do not evolve in fixed steps.

Theorem 32.1

1. If a nonexplosive continuous-time Markov chain is irreducible and positive recurrent, then there exists a unique vector

   $$\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \ldots),$$

   with nonnegative entries summing to 1, such that

   $$\boldsymbol{\pi}\,\mathbf{P}(t) = \boldsymbol{\pi} \quad \text{for all } t \in \mathbb{R}^+ \cup \{0\}.$$

Furthermore,

$$\lim_{t\to\infty} \mathbf{P}(t) = \mathbf{\Pi} = \begin{bmatrix} \pi_0 & \pi_1 & \cdots \\ \pi_0 & \pi_1 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Here π_i is the proportion of time the chain spends in state i. If $\nu_i > 0$ is the rate of leaving i and T_{ii}^+ is the elapsed time from an entrance into i to the next entrance, then

$$\mathbb{E}_i[T_{ii}^+] = \frac{1}{\pi_i \nu_i}.$$

2. In any stationary probability distribution, every transient or null recurrent state has mass $\pi_i = 0$.

Example 32.2 For $S = \{0, 1\}$, given

$$\mathbf{P}(t) = \begin{bmatrix} \mathbb{P}_{00}(t) & \mathbb{P}_{01}(t) \\ \mathbb{P}_{10}(t) & \mathbb{P}_{11}(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{3} + \frac{2}{3}e^{-6t} & \frac{2}{3} - \frac{2}{3}e^{-6t} \\ \frac{1}{3} - \frac{1}{3}e^{-6t} & \frac{2}{3} + \frac{1}{3}e^{-6t} \end{bmatrix},$$

verify the transition matrix is valid, classify the chain, and determine the stationary distribution.

Solution. First we verify that $\mathbf{P}(t)$ is valid. We have

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

For any t, the matrix is square, its entries are nonnegative, and its rows sum to 1. To check the semigroup property without suppressing the calculation, write

$$\mathbf{\Pi} = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix} \qquad \text{and} \qquad \mathbf{P}(t) = \mathbf{\Pi} + e^{-6t}(\mathbb{I} - \mathbf{\Pi}).$$

Since $\mathbf{\Pi}^2 = \mathbf{\Pi}$ and $\mathbf{\Pi}(\mathbb{I} - \mathbf{\Pi}) = (\mathbb{I} - \mathbf{\Pi})\mathbf{\Pi} = \mathbf{0}$, multiplication gives $\mathbf{P}(t)\mathbf{P}(s) = \mathbf{P}(t+s)$. Thus the Chapman–Kolmogorov equations hold. Now we classify the chain. We have $0 \to 1$ since $\mathbb{P}_{01}(t) > 0$ and $1 \to 0$ since $\mathbb{P}_{10}(t) > 0$ for every $t > 0$, so it is irreducible and positive recurrent by finiteness. The embedded jump chain alternates deterministically, so its transition matrix is

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

Finally, in the long run,

$$\lim_{t\to\infty} \mathbf{P}(t) = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix},$$

so

$$\boldsymbol{\pi} = \left(\frac{1}{3}, \frac{2}{3}\right).$$

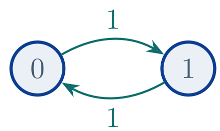

FIGURE 14

Figure 14 shows its embedded jump chain: both transition probabilities are one. The continuous-time jump rates are 4 from state 0 and 2 from state 1.

□

In that example, we started with $\mathbf{P}(t)$; in practice, we more often start with holding-time distributions and the embedded transition matrix. We therefore want one matrix that combines both pieces of information.

We want a single matrix that is easy to obtain, gives us $\boldsymbol{\pi}$, and describes short-term behaviour in genuinely continuous time.

Lecture 33 — Wednesday, July 23, 2014

The Generator Q

Definition 33.1 For the regular continuous-time Markov chains considered here, the transition functions $\mathbb{P}_{ij}(t)$ are right-differentiable at zero. We define

$$\mathbf{Q} = \mathbf{P}'(0^+) = \lim_{h\to 0^+} \frac{\mathbf{P}(h) - \mathbf{P}(0)}{h} = \lim_{h\to 0^+} \frac{\mathbf{P}(h) - \mathbb{I}}{h},$$

where the derivative is taken componentwise. That is $\mathbf{Q} = (q_{ij})_{i,j\in S}$ where

$$q_{ij} = \lim_{h\to 0^+} \frac{\mathbb{P}_{ij}(h)}{h} \text{ for all } i \neq j,$$

and

$$q_{ii} = \lim_{h\to 0^+} \frac{\mathbb{P}_{ii}(h) - 1}{h}.$$

$\mathbf{Q}$ is called a **generator** (or **infinitesimal generator**).

The generator is sometimes denoted R, G, A, or X. Under the standard nonexplosiveness assumptions used here, it contains all information about the chain, encoding jump times, possible transitions of the embedded chain ($\mathbf{P}$), $\boldsymbol{\pi}$, $\mathbf{P}(t)$, and related quantities.

Proposition 33.2 Let $\mathbf{Q} = (q_{ij})$ be the generator of a finite-state continuous-time Markov chain. More generally, assume the limits below may be interchanged with the row sum. Then:

1. $q_{ij} \geq 0$ for all $i \neq j$.
2. $\sum_{j\in S} q_{ij} = 0$ for each $i \in S$.

Consequently, $q_{ii} = -\sum_{j\neq i} q_{ij} \leq 0$, so $\mathbf{Q}$ is not a stochastic matrix.

Proof. For $i \neq j$,

$$q_{ij} = \lim_{h\to 0^+} \frac{\mathbb{P}_{ij}(h)}{h} \geq 0,$$

because $\mathbb{P}_{ij}(h) \geq 0$ and $h > 0$.

Next, fix $i \in S$. Using the definition of the diagonal and off-diagonal entries,

$$\sum_{j\in S} q_{ij} = \sum_{j\neq i} \lim_{h\to 0^+} \frac{\mathbb{P}_{ij}(h)}{h} + \lim_{h\to 0^+} \frac{\mathbb{P}_{ii}(h) - 1}{h} = \lim_{h\to 0^+} \frac{\sum_{j\in S} \mathbb{P}_{ij}(h) - 1}{h}.$$

Since each row of $\mathbf{P}(h)$ sums to 1, we have $\sum_{j\in S} \mathbb{P}_{ij}(h) = 1$, so

$$\sum_{j\in S} q_{ij} = \lim_{h\to 0^+} \frac{1 - 1}{h} = 0.$$

Therefore $q_{ii} = -\sum_{j \neq i} q_{ij} \leq 0$. □

Example 33.3 Consider the setup from Example 32.2: $S = \{0, 1\}$ and

$$\mathbf{P}(t) = \begin{bmatrix} \frac{1}{3} + \frac{2}{3}e^{-6t} & \frac{2}{3} - \frac{2}{3}e^{-6t} \\ \frac{1}{3} - \frac{1}{3}e^{-6t} & \frac{2}{3} + \frac{1}{3}e^{-6t} \end{bmatrix}.$$

Then

$$\mathbf{Q} = \mathbf{P}'(0^+) = \begin{bmatrix} -4e^{-6t} & 4e^{-6t} \\ 2e^{-6t} & -2e^{-6t} \end{bmatrix}_{t=0^+} = \begin{bmatrix} -4 & 4 \\ 2 & -2 \end{bmatrix}.$$

So $q_{ij} \geq 0$ for $i \neq j$, and the rows sum to 0. ✓

However, we don't usually start with $\mathbf{P}(t)$. We want to get $\mathbf{Q}$ directly from the description of the Markov chain. Recall the standard continuous-time Markov chain construction facts:

1. The amount of time spent in each nonabsorbing state i before jumping (call it L_i) has the memoryless property. Thus $L_i \sim \text{Exp}(\nu_i)$ for some $\nu_i > 0$. If $\nu_i = 0$, state i is absorbing and $L_i = \infty$.
2. The embedded chain probabilities $(\mathbb{P}_{ij})$ determine where we jump when leaving i. So

$$\begin{aligned} \nu_i \mathbb{P}_{ij} &= (\text{rate leaving } i)\mathbb{P}(\text{go to } j \mid \text{leaving } i) \\ &= \text{instantaneous transition rate from } i \text{ to } j, \qquad j \neq i. \end{aligned}$$

Finding Q Directly

Proposition 33.4 $q_{ii} = -\nu_i$ and $q_{ij} = \nu_i \mathbb{P}_{ij}$ for $i \neq j$.

Proof. Note that with exponential waiting times, the probability of 2 or more jumps in a small interval h is $o(h)$, and the probability of 1 jump in a small interval h is $\nu_i h + o(h)$. Therefore,

$$\begin{aligned} q_{ii} &= \lim_{h \to 0^+} \frac{\mathbb{P}_{ii}(h) - 1}{h} = \lim_{h \to 0^+} \frac{\mathbb{P}(\text{stay in } i) + \mathbb{P}(\text{go away and come back}) - 1}{h} \\ &= \lim_{h \to 0^+} \frac{[e^{-\nu_i h} + o(h)] - 1}{h} = \lim_{h \to 0^+} \left[\frac{e^{-\nu_i h} - 1}{h} + \frac{o(h)}{h} \right] = \frac{\mathrm{d}}{\mathrm{d}t} e^{-\nu_i t} \bigg|_{t=0} = -\nu_i \end{aligned}$$

and

$$\begin{aligned}
q_{ij} = \lim_{h\to 0^+} \frac{\mathbb{P}_{ij}(h)}{h} &= \lim_{h\to 0^+} \frac{\mathbb{P}(\text{one jump from } i \text{ to } j) + \mathbb{P}(\text{somewhere else, then to } j)}{h} \\
&= \lim_{h\to 0^+} \frac{\mathbb{P}(\text{one jump from } i)\mathbb{P}(\text{to } j \mid \text{from } i) + o(h)}{h} \\
&= \lim_{h\to 0^+} \frac{[\nu_i h + o(h)]\mathbb{P}_{ij} + o(h)}{h} = \nu_i \mathbb{P}_{ij}. \qquad \square
\end{aligned}$$

So if we know the rates of each state and the embedded chain probabilities, we obtain $\mathbf{Q}$. What can we do with $\mathbf{Q}$?

1. For a finite-state chain, we can use it to solve for $\boldsymbol{\pi}$ by finding a nonnegative vector such that $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, with the π_i summing to 1.
2. We can find the ν_i and, when $\nu_i > 0$, the $\mathbb{P}_{ij}$. We know that $q_{ii} = -\nu_i$, so $\nu_i = |q_{ii}|$, and

$$\mathbb{P}_{ij} = \frac{q_{ij}}{\nu_i} = \frac{q_{ij}}{|q_{ii}|}.$$

Lecture 34 — Thursday, July 24, 2014

Birth and Death Processes

Let us construct a continuous-time Markov chain on $S = \{0, 1, 2, \dots\}$ or $\{0, 1, \dots, k\}$: a so-called **birth and death process**. Each transition must be either a "birth" or "death"; for example, $X(t)$ may denote population size at time t, or the number of people in line. The transition rates out of state i can depend on i, but not on the history of $X(t)$.

If $X(t) = i$, let λ_i = birth rate and μ_i = death rate, with nonnegative rates and independent clocks. At the lower boundary $\mu_0 = 0$, and for a finite upper boundary k, $\lambda_k = 0$. Think of $T_B \sim \text{Exp}(\lambda_i)$ as the "birth clock" and $T_D \sim \text{Exp}(\mu_i)$ as the "death clock" whenever the corresponding rate is positive.

We can find the total rate out of each state i as follows. If $i = 0$, then $L_0 = T_B \sim \text{Exp}(\lambda_0)$, so $\nu_0 = \lambda_0$. For $i > 0$, $L_i = \min(T_B, T_D) \sim \text{Exp}(\lambda_i + \mu_i)$, so $\nu_i = \lambda_i + \mu_i$.

We can also find the embedded chain probabilities:

$$\mathbb{P}_{0,1} = \mathbb{P}(\text{go to 1} \mid \text{leaving 0}) = 1,$$
$$\mathbb{P}_{i,i+1} = \mathbb{P}(T_B < T_D) = \frac{\lambda_i}{\lambda_i + \mu_i} \qquad \text{and} \qquad \mathbb{P}_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}.$$

All other $\mathbb{P}_{ij}$ are 0. So we can get the matrix $\mathbf{Q}$ using $q_{ii} = -\nu_i$ and $q_{ij} = \nu_i \mathbb{P}_{ij}$.

For example, if $k = 3$ is the maximum,

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 \\ 0 & 0 & \mu_3 & -\mu_3 \end{bmatrix}.$$

Assume the finite birth and death chain is irreducible, so the adjacent rates used below are positive. We can now solve for $\boldsymbol{\pi}$ using $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$. Since $\mathbf{Q}$ is relatively simple, we can solve this system easily: we have

$$\pi_0(-\lambda_0) + \pi_1(\mu_1) = 0 \Rightarrow \pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$$

and

$$\pi_0(\lambda_0) + \pi_1(-(\lambda_1 + \mu_1)) + \pi_2(\mu_2) = 0 \Rightarrow \pi_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0.$$

So

$$\pi_3 = \frac{\lambda_2}{\mu_3}\pi_2 = \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}\pi_0.$$

Hence

$$\boldsymbol{\pi} = \left(\pi_0, \frac{\lambda_0}{\mu_1}\pi_0, \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0, \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}\pi_0\right),$$

with

$$\pi_0 = \left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}\right)^{-1}.$$

Lecture 35 — Monday, July 28, 2014

Example 35.1 Suppose $S = \{0, 1, 2\}$ and

$$\mathbf{Q} = \begin{bmatrix} -2/5 & 1/5 & 1/5 \\ 0 & -3 & 3 \\ 1/4 & 3/4 & -1 \end{bmatrix}.$$

Find the distribution of time spent in each state, the embedded chain, and the equilibrium distribution.

Solution. We have $L_0 \sim \text{Exp}(2/5)$, so the average time is $5/2 = 2.5$ minutes. Similarly, $L_1 \sim \text{Exp}(3)$ with average time $1/3$ minutes, and $L_2 \sim \text{Exp}(1)$ with average time 1 minute. We can also find the embedded chain: $\mathbb{P}_{01} = \mathbb{P}_{02} = 1/2$ since $q_{01} = q_{02}$, and

$$\mathbb{P}_{10} = 0, \qquad \mathbb{P}_{12} = 1, \qquad \mathbb{P}_{20} = \frac{1}{4}, \qquad \text{and} \qquad \mathbb{P}_{21} = \frac{3}{4}.$$

Thus

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

Note that the embedded chain has a zero diagonal because none of these states is absorbing.

The corresponding embedded jump chain is shown in Figure 15.

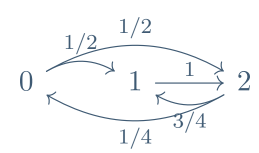

Figure 15

There is no periodicity in continuous time. Finally, we find the equilibrium distribution by solving $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, subject to $\pi_0 + \pi_1 + \pi_2 = 1$, which gives

$$\boldsymbol{\pi} = \left(\frac{15}{46}, \frac{7}{46}, \frac{24}{46} \right).$$

□

How can we find $\mathbf{P}(t)$ using $\mathbf{Q}$? We use the Kolmogorov forward and backward equations.

Definition 35.2 For a finite-state continuous-time Markov chain (and, more generally, under the regularity conditions that justify the sums), in matrix form,

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

is the **Kolmogorov forward equation** and

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

is the **Kolmogorov backward equation**, with boundary condition $\mathbf{P}(0) = \mathbb{I}$. In pointwise form,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}_{ij}(t) = \sum_{k\in S}\mathbb{P}_{ik}(t)\cdot q_{kj} = \sum_{k\in S} q_{ik}\cdot\mathbb{P}_{kj}(t).$$

Example 35.3 Let $S = \{0, 1\}$ and

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

for $\lambda, \mu > 0$ (this is the irreducible birth and death process with two states). We know that

$$\mathbf{P}(t) = \begin{bmatrix} \mathbb{P}_{00}(t) & 1 - \mathbb{P}_{00}(t) \\ 1 - \mathbb{P}_{11}(t) & \mathbb{P}_{11}(t) \end{bmatrix}.$$

By the Kolmogorov forward equation, $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$, so

$$\mathbb{P}'_{00}(t) = -\lambda\mathbb{P}_{00}(t) + \mu(1 - \mathbb{P}_{00}(t)) = -(\lambda+\mu)\mathbb{P}_{00}(t) + \mu.$$

The solution to this differential equation is

$$\mathbb{P}_{00}(t) = e^{-(\lambda+\mu)t}\left(C + \frac{\mu}{\lambda+\mu}e^{(\lambda+\mu)t}\right) = \frac{\mu}{\lambda+\mu} + Ce^{-(\lambda+\mu)t}.$$

Since $\mathbb{P}_{00}(0) = 1$, we have $C = \frac{\lambda}{\lambda+\mu}$. Therefore

$$\mathbb{P}_{00}(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}.$$

The other $\mathbb{P}_{ij}(t)$ values then follow, and we recover the same $\mathbf{P}(t)$ matrix as in the earlier case with $\lambda = 4, \mu = 2$.

There are many examples of birth and death processes. For example:

1. The Poisson process (a pure birth process with rate λ).
2. A single-server queue: Its transition structure is shown in Figure 16.

$$\begin{tikzcd} 0 \arrow[r, bend left, "\lambda"] & 1 \arrow[l, bend left, "\mu"] \arrow[r, bend left, "\lambda"] & 2 \arrow[l, bend left, "\mu"] \arrow[r, bend left, "\lambda"] & \cdots \end{tikzcd}$$

FIGURE 16

Here λ is the arrival rate, so $\lambda_i = \lambda$ for all i, and μ is the service rate, so $\mu_i = \mu$ for all $i \geq 1$. If $\rho = \lambda/\mu < 1$, then the equilibrium distribution is geometric:

$$\pi_i = (1-\rho)\rho^i, \qquad i = 0, 1, 2, \ldots.$$

3. Consider three machines and two mechanics. The breakdown time has mean 10 weeks per machine, and repair time has mean 8 weeks. Let $X(t)$ be the number of machines working at time t, so $S = \{0, 1, 2, 3\}$. The rates in Figure 17 account for both available mechanics and all machines that can fail.

$$\begin{tikzcd} 0 \arrow[r, bend left, "2/8"] & 1 \arrow[l, bend left, "1/10"] \arrow[r, bend left, "2/8"] & 2 \arrow[l, bend left, "2/10"] \arrow[r, bend left, "1/8"] & 3 \arrow[l, bend left, "3/10"] \end{tikzcd}$$

FIGURE 17

$$\pi_0 = \left[1 + \frac{2/8}{1/10} + \frac{(2/8)(2/8)}{(1/10)(2/10)} + \frac{(2/8)(2/8)(1/8)}{(1/10)(2/10)(3/10)}\right]^{-1} \approx 0.1261\ldots$$

and

$$\pi_1 = \frac{2/8}{1/10}\pi_0, \qquad \pi_2 = \frac{2/8}{2/10}\pi_1, \qquad \text{and} \qquad \pi_3 = \frac{1/8}{3/10}\pi_2.$$

We can also find $\mathbb{E}[X(t)] = 0 \cdot \pi_0 + 1 \cdot \pi_1 + 2 \cdot \pi_2 + 3 \cdot \pi_3 = 1.597\ldots$ The proportion of time that both mechanics work in the long run is $\pi_0 + \pi_1$.

4. A linear/population growth model. Suppose we have a bacteria population that reproduces by splitting into two bacteria at rate λ. Each bacterium dies independently at rate μ. Let $X(t)$ be the number of bacteria alive at time t. Then $S = \{0, 1, 2, \dots\}$. Figure 18 shows how both rates scale with the population.

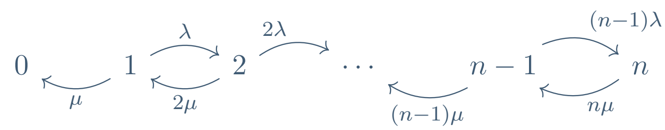

FIGURE 18

In general, $\mu_i = i \cdot \mu$ and $\lambda_i = i \cdot \lambda$, so the rates are linear in the state.

Appendix: Lecture and Tutorial Index