STAT 380 Week 1

Course outline

Reading : Chapters 1, 2 and 3 of Ross.
Goals for the Week:

- Give an overview of the course.
- Do some review by example.

Course outline:

- Basic concepts of probability. Review of
  - Distributions
  - Expectation and moments
  - Moment generating functions
  - Independence, conditioning

  - Markov Chains
  - Poisson Processes
  - Birth and Death Processes
  - Brownian Motion
  - Monte Carlo Simulation techniques
    * Random number generators
    * Generating random variables

Summary

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

– Baye's rule

- Introduction to modelling

Basic Examples

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

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

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

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

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

All of the A_i are equally likely.

Apply two rules:

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

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

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

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

Definition of conditional probability:

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

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

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

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

Model? Traditionally we define events like

$$A_i = \{\text{Prize behind door } i\}$$
$$B_j = \{\text{You choose door } j\}$$

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

The event LS, that you lose if you switch is

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

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

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

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

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

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

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

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

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

1. Monte picks at random so

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

2. Monte chooses the door with the largest possible number:

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

Use *Bayes' rule*:

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

Numerator is

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

Denominator is

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

which simplifies to

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

which in turn is

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

For case 1 we get

$$P(M=1|C=1,S=3)=\frac{1/2}{1/2+1}=\frac{1}{3}$$

while for case 2 we get

$$P(M=1|C=1,S=3)=\frac{1}{1+1}=\frac{1}{2}$$

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

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

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

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

Useful rvs:

$$X=\#\text{ of male children of first man}$$

$$Z_k=\#\text{ of male children in generation }k$$

Event of interest:

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

Break up E_n:

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

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

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

Then

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

Now add modelling assumptions:

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

2. Given $X = k$ the probability of $B_{j,n-1}$ is q_{n-1}. In other words: sons are just like the parent.

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

$$\begin{aligned} q_n &= \sum_{k=0}^{\infty} P(E_n \cap \{X = k\}) \\ &= \sum_{k=0}^{\infty} P(\cap_{j=1}^k B_{j,n-1} | X = k) p_k \\ &= \sum_{k=0}^{\infty} \prod_1^k P(B_{j,n-1} | X = k) p_k \\ &= \sum_{k=0}^{\infty} (q_{n-1})^k p_k \end{aligned}$$

Probability generating function:

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

We have found

$$q_1 = p_0$$

and

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

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

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

Special cases

Geometric Distribution: Assume

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

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

Then

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

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

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

Two roots are

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

One of the roots is 1; the other is

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

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

Binomial(m, θ): If

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

then

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

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

Poisson(λ): Now

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

and

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

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

Important Points:

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