MATH 2030 3.00 – Elementary Probability
Course Notes
Part I: Models and Counting

Tom Salisbury – salt@yorku.ca

York University, Dept. of Mathematics and Statistics
Original version April 2010. Thanks are due to
E. Brettler, V. Michkine, and R. Shieh for many corrections

January 18, 2018

Introduction

- *Probability*: the mathematics used for Statistics
  Are related but different.
  - Probability question: Assume a coin is fair. Flip it 50 times. What is the probability of getting 28 heads?
  - Statistics question: Flip a coin 50 times and observe 28 heads. Do you believe it is fair?
- Also fundamental in: mathematical modelling, modern physics (statistical mechanics), electrical engineering (signal noise), computer science (simulation), finance (option pricing), etc.
- *Frequentist interpration*: repeat an experiment a large number of times. The probability of an event is the relative frequency with which it occurs, in the long run.
- Other interpretations: level of uncertainty, or subjective belief.
- First task: build a mathematical model that captures our empirical sense of probabilities.

Model

Model has three ingredients:

- a non-empty set Ω
- a collection $\mathcal{F}$ of subsets of Ω, satisfying
  - $\emptyset \in \mathcal{F}$
  - $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
  - $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \cup A_i \in \mathcal{F}$
- a function $P : \mathcal{F} \to \mathbb{R}$ satisfying
  - $P(A) \geq 0 \quad \forall A \in \mathcal{F}$
  - $P(\Omega) = 1$
  - $A_1, A_2, \dots \in \mathcal{F}$ and disjoint $\Rightarrow P(\cup A_i) = \sum P(A_i)$.

Remarks: Here $A^c = \Omega \setminus A$, where $B \setminus A = \{\omega \in B \mid \omega \notin A\}$. The model is $(\Omega, \mathcal{F}, P)$. $\mathcal{F}$ as above is called a *σ-field*, and P a *probability measure*. The sequence $A_1, A_2, \dots$ above can be finite or countable, and the last property is called *countable additivity*.

Interpretation

- Elements $\omega \in \Omega$ are called *outcomes* (or *sample points* or *states of nature*).
- Ω is the *sample space*, ie. the set of all possible outcomes. We think of the "experiment" as choosing ω at random from Ω. So knowing what ω is chosen determines everything else: ω should code within it all information we care about, and answers to all questions we will ask.
- *Events* A are elements of $\mathcal{F}$, so $A \subset \Omega$. If ω is the outcome "picked", then we say A *occurs* if $\omega \in A$.
- $P(A)$ is the *probability* that the event A occurs. $\mathcal{F}$ is the set of events to which we can assign probabilities.
- *Random variables* are measurements, whose values depend on ω. Formally, are functions $X : \Omega \to \mathbb{R}$ such that

$$\{\omega \mid X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

Example

Flip a fair coin twice:

- eg. if the first flip is "head" and the second is "tails" we represent this by $\omega = (H, T)$.
- So $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$ has 4 elements.
- We'll assign prob's to all subsets of Ω. So $\mathcal{F}$ consists of all $16 = 2^4$ subsets of Ω.
- eg. the event that the flips agree is $\{(H,H),(T,T)\}$, and contains 2 outcomes.
- "Fair" means all 4 outcomes are equally likely.
  So let $P(A) = \#(A)/4$.
  eg. prob of 2 heads is $P(\{(H,H)\}) = \frac{1}{4}$.
  eg. prob that flips agree is $P(\{(H,H),(T,T)\}) = \frac{1}{2}$.

Example (continued)

- Let the random variable X be the number of heads.
  - $X((H,H)) = 2$
  - $X((H,T)) = 1 = X((T,H))$
  - $X((T,T)) = 0$.

  So $P(\{\omega \mid X(\omega) = 2\}) = \frac{1}{4}$, $P(\{\omega \mid X(\omega) = 1\}) = \frac{1}{2}$, $P(\{\omega \mid X(\omega) = 0\}) = \frac{1}{4}$.

  Remark: We abbreviate these as $P(X = 2)$, $P(X = 1)$, and $P(X = 0)$. More generally, for any real number x we write $P(X = x)$ as an abbreviation for $P(\{\omega \mid X(\omega) = x\})$.
- There are 16 events in $\mathcal{F}$:
  - 1 with 0 outcomes, namely $\emptyset$
  - 4 with 1 outcome, eg $\{(H,H)\}$ (both flips are heads)
  - 6 with 2 outcomes, eg $\{(H,H),(H,T)\}$ (first flip is a head)
  - 4 with 3 outcomes, eg $\{(H,H),(H,T),(T,H)\}$ (at least one is a head)
  - 1 with 4 outcomes, namely Ω itself.

Example (continued)

In fact, to write it all out:

$$\begin{aligned}\mathcal{F} = \Big\{&\emptyset, \{(H,H)\}, \{(H,T)\}, \{(T,H)\}, \{(T,T)\},\\ &\{(H,H),(H,T)\}, \{(H,H),(T,H)\}, \{(H,H),(T,T)\},\\ &\{(H,T),(T,H)\}, \{(H,T),(T,T)\}, \{(T,H),(T,T)\},\\ &\{(H,H),(H,T),(T,H)\}, \{(H,H),(H,T),(T,T)\},\\ &\{(H,H),(T,H),(T,T)\}, \{(H,T),(T,H),(T,T)\},\\ &\{(H,H),(H,T),(T,H),(T,T)\}\Big\}.\end{aligned}$$

- To model flipping unfair (or linked) coins, keep same $\mathcal{F}$ but change P.

Interpretation

- A^c is the complimentary event to A. A^c occurs $\Leftrightarrow$ A doesn't.
- $\emptyset$ never occurs. A null event.
- Ω always occurs. A sure event.
- $A \cup B$ occurs $\Leftrightarrow$ A occurs or B occurs (ie if at least one does)
- $A \cap B$ occurs $\Leftrightarrow$ both A and B occur.
- A and B are disjoint $\Leftrightarrow$ mutually exclusive
- $A \subset B$ means that $\omega \in A \Rightarrow \omega \in B$.
  That is, if A occurs, then B occurs too.
- For discrete Ω (ie finite or countable), can take $\mathcal{F}$ to consist of **ALL** subsets of Ω. And **ALL** $X : \Omega \to \mathbb{R}$ as r.v. This won't work in general, but in this course we will behave as though this is always the case. ie we will mostly ignore $\mathcal{F}$ now.

Models

- Whether we can use simple models depends on the questions we'll ask, since ω has to code all the answers. Eg to model stock market for a month would take each ω to be a list of ALL prices at ALL times. Eg a full statistical mechanics model takes ω to code positions and momenta of ALL particles. Often we need to know that a model $\exists$, but don't want to actually write it down.
- Have a choice of many ways to model the same experiment. In fact (as soon as we have enough foundation) will try to do higher-level calculations without actually specifying the details of the model we're using.

Properties / Consequences of the axioms

- $P(A^c) = 1 - P(A)$ (as $A \cup A^c = \Omega$ disjointly).
- $0 \leq P(A) \leq 1$ (as $P(A^c) \geq 0$).
- $A \subset B \Rightarrow P(A) \leq P(B)$ (as $B = A \cup (B \setminus A)$ disjointly).
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  (pf: break up disjointly as $(A \setminus B) \cup (A \cap B) \cup (B \setminus A)$.
  Then use additivity.)
- $$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$
  This is called *inclusion/exclusion*.
  Proof can be given as above, or using *Venn diagrams*.

Discrete Models

- A model is *discrete* if $\Omega = \{\omega_1, \omega_2, \dots\}$ is finite or countable.
- In that case, P is determined by fixing numbers $p_1, p_2, \dots$ such that $0 \leq p_i \leq 1 \; \forall i$, and $\sum p_i = 1$. Just take

$$P(A) = \sum_{i \text{ s.t. } \omega_i \in A} p_i.$$

- Special case: **equally likely outcome models**
  Ω finite, all p_i equal. Then $p_i = 1/\#(\Omega)$, and

$$P(A) = \frac{\#(A)}{\#(\Omega}.$$

In this case we will calculate by counting.

More complicated models

- Not all models are discrete.

Example: *Model for a uniform r.v..*
$\Omega = [0,1)$, $P(A) = |A|$. ($\mathcal{F}$ is complicated. = "Borel sets". This is a case where it CAN'T be all subsets, only reasonable ones).
Take $Y(\omega) = \omega \in \mathbb{R}$.
eg. $P(Y \leq x) = |[0,x]| = x$, for $0 \leq x < 1$.
We'll study such r.v. later (using calculus).

Example: *Model for ∞ many coin flips.*
$\Omega = [0,1)$, $P(A) = |A|$ (as before)
Write $\omega = 0.\omega_1\omega_2\omega_3\ldots$ in binary (eg $\frac{1}{2} = .1$, $\frac{3}{4} = .11$, $\frac{1}{4} = .01$).
If 0 represents T and 1 represents H, then the ith coin flip r.v. is $X_i(\omega) = \omega_i$.
eg. $P(X_1 = 0) = |[0, \frac{1}{2})| = \frac{1}{2}$, $P(X_2 = 0) = |[0, \frac{1}{4}) \cup [\frac{1}{2}, \frac{3}{4})| = \frac{1}{2}$.
$P(X_1 = 0 \text{ and } X_2 = 1) = |[\frac{1}{4}, \frac{1}{2})| = \frac{1}{4}$.

Discrete r.v.

- A r.v. has a *discrete distribution* if there are only finitely or countably many values it can take.
- In that case, the distribution is given by a table of values x and probabilities $P(X = x)$. Eg. No. of H in 2 coin flips:

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Eg. Roll 2 dice, $X =$ sum of the numbers showing.
Take an equally likely outcome model,
$\Omega = \{(i, j) \mid 1 \leq i, j \leq 6\} = \{(1, 1), \ldots, (6, 6)\}$.
So $\{\omega \mid X(\omega) = 2\} = \{(1, 1)\}$ has probability $\frac{1}{36}$,
$\{\omega \mid X(\omega) = 3\} = \{(1, 2), (2, 1)\}$ has probability $\frac{2}{36}$.

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Counting

Want to count $\#(A)$, the number of elements in a set A.
Do this by finding a list of k properties that each element has.
Need to do this so the following hold

- Each property can be specified some fixed number of ways, *regardless of how the other properties were specified*.
  Let n_i be the number of choices for the ith property.
- There is a one-to-one correspondence between ways of specifying all these properties, and elements of A.

Then $\#(A) = n_1 \times n_2 \times \cdots \times n_k$. **[Basic counting principle]**

- **Eg. Dice:** We just saw that there were 36 ways to roll two dice in sequence. In this framework, there are two properties in our list: the first roll (6 possibilities), and the second roll (6 possibilities). $36 = 6 \times 6$.

Counting

- **Subsets:** If A has n elements, how many subsets $B \subset A$ are there? Here there are n properties in our list: Is the 1st element in or out? Is the 2nd element in or out? etc. There are two possibilities for each, so there are $2 \times 2 \times \cdots \times 2 = 2^n$ subsets.
- **Eg. Cards:** Each card has two properties: a *value* and a *suit*. There are 13 ways of specifying the value (Ace, 2, 3, 4, 5, 6, 7, 8, 9, Jack, Queen, King) and 4 ways of specifying the suit (clubs ♣ , diamonds ♢, hearts ♡, spades ♠). Specifying a card is the same as specifying its value and suit, so $\#(\text{cards}) = 13 \times 4 = 52$.
  [In future, spades and clubs will be called *black* suits, and diamonds and hearts will be called *red* suits]

Eg. 3-card hands

Deal out 3 cards in sequence, without replacement. ie a card hand is (a, b, c) where a is the 1st card dealt, b is the 2nd card dealt, and c is the 3rd card dealt. Want to count the number of:

- **3-card hands:** Imagine ranking the 52 cards in some way. The properties are: the first card (52 ways); the ranking of the 2nd card among the remaining cards (51 ways); the ranking of the 3rd card among the remaining cards (50 ways). so $\#(\text{hands}) = 52 \times 51 \times 50$.
- **3-card hands with same value:** $52 \times 3 \times 2$
- **3-card hands with same suit:** $52 \times 12 \times 11$
- **3-card hands with 1 pair:** (ie 2 with $=$ value, other $\neq$). Properties: value for pair, 1st pair suit, 2nd pair suit, other card, deal for other card: $13 \times 4 \times 3 \times 48 \times 3$
  [Bad properties: 1st card (52 ways), 2nd (51 ways), 3rd card (48 or 6 ways, depending on the 1st and 2nd choices). So would have to break up as $52 \times 3 \times 48 + 52 \times 48 \times 6$]

Permutations

- Choose k **distinct** objects from n objects, **in order**. In other words, let $\#(A) = n$, and take

$$(n)_k = \#\{(a_1, a_2, \dots, a_k) \mid \text{each } a_i \in A, \text{ all } a_i \text{ distinct}\},$$

the number of *permutations of k objects from n objects*.

- So $(n)_k = \underbrace{n(n-1)(n-2)\dots(n-k+1)}_{k} = \dfrac{n!}{(n-k)!}$, where $n! = n(n-1)\cdots(3)(2)(1)$.
[Sometimes people write ${}_nP_k$ for $(n)_k$.]
- **Eg:** The number of ways of *ordering* n objects is $(n)_n = n!$
- **Eg:** The number of 3-card hands (dealt in order) is $(52)_3 = 52 \times 51 \times 50$.
- Convention: $0! = 1$ and $(n)_0 = 1$.

Combinations

- Choose k **distinct** objects from n objects, **without order**. In other words, let $\#(A) = n$, and take

$$\binom{n}{k} = \#\{B \mid B \text{ is a } k\text{-element subset of } A\},$$

the number of *combinations of k objects from n objects*. This is read "n choose k" [and sometimes people write ${}_nC_k$ for $\binom{n}{k}$.] Also called a *binomial coefficient*.

- If we count permutations by counting first the subset, and then the way it is ordered, we get $(n)_k = \binom{n}{k} \cdot k!$, that is,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- So $\binom{n}{n} = 1 = \binom{n}{0}$, $\binom{n}{1} = n$, $\binom{n}{k} = \binom{n}{n-k}$.

Poker hands without order

Deal out 5 cards without replacement, and without distinguishing on the basis of the sequence they're dealt in. ie a card hand is a 5-card subset of the set of 52 cards. Want to count the number of:

- **hands:** $\binom{52}{5} = 2,598,960$
- **flushes:** $\binom{4}{1}\binom{13}{5} = 5,148$. [Really should subtract the probability of a *straight flush*].
- **3 of a kind:** $\binom{13}{1}\binom{12}{2}\binom{4}{3}\binom{4}{1}\binom{4}{1}$ [Choose 3-value, other values, 3-suits, suit for lowest other value, suit for highest other value. Note that we are excluding a *full house* or a *4 of a kind*]
- **pairs:** $\binom{13}{1}\binom{12}{3}\binom{4}{2}\binom{4}{1}\binom{4}{1}\binom{4}{1}$.
- In a fair deal, all outcomes are equally likely. Can work out probabilities either with order (permutations) or without (combinations). Counting is different, but probabilities should be the same.

Binomial theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

- Expand $(a+b)^n = (a+b)(a+b)\cdots(a+b)$. Get a sum of terms $c_1 c_2 \cdots c_n$ where each c_i is a or b. How many are $a^k b^{n-k}$? Choose the k locations corresponding to a's, so $\binom{n}{k}$.
- Eg. $(a+b)^4 = a^4 + 4a^3 b + 6a^2 b^2 + 4ab^3 + b^4$.
- Recurrence $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.
  [k element subsets of $\{1, 2, \ldots, n\}$ either include n or they don't. $\binom{n-1}{k-1}$ do, and $\binom{n-1}{k}$ don't]
- This $\Rightarrow$ *Pascal's triangle* (see text). An inefficient way of finding one binomial coefficient, but a good way of finding a whole row.

Binomial distribution

- Let $0 \le p \le 1$. Say that a random variable X has a *binomial distribution* with parameters n and p, or $X \sim \text{Bin}(n, p)$, if

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

- This makes sense, because all these are ≥ 0, and by the binomial theorem, they sum to $\Big(p + (1-p)\Big)^n = 1^n = 1$.
- We'll study this distribution later, where it will be important for things like counting the number of successes in independent trials.

Sampling

- **With replacement:** An urn has 12 red balls and 8 green balls. Pick 5, with replacement. $P(3R, 2G) =$?? Take Ω = all sequences $(a_1, a_2, \dots, a_5)$ chosen from $\{R_1, R_2, \dots, R_{12}, B_1, \dots, B_8\}$. $\#(\Omega) = 20^5$. The event A has $\#(A) = \binom{5}{3}12^3 8^2$. So $P(A) = \binom{5}{3}(\frac{12}{20})^3(\frac{8}{20})^2 \approx 0.3456$ [In fact, the number of reds has a binomial distribution.]
- **Without replacement:** Now Ω = all 5-element subsets of $\{R_1, R_2, \dots, R_{12}, B_1, \dots, B_8\}$. $\#(\Omega) = \binom{20}{5} = 15,504$ and $\#(A) = \binom{12}{3}\binom{8}{2} = 6,160$, so

$$P(A) = \frac{\binom{12}{3}\binom{8}{2}}{\binom{20}{5}} \approx 0.3456$$

[Now the number of reds has what is called a hypergeometric distribution]
- Either case could be done with or without order, but these choices are easiest.