Review of Lecture #3

- We explored the consequences of the axioms of probability
- We discussed the notion of a partition and how partitions can be used to calculate the probability of an event
- We discussed binomial coefficients
- We looked at a simple combinatorial problem

Drawing a random sample

- Example: An urn contains 6 identical red balls R_1, R_2, R_3, R_5, R_5, R_6 and 3 identical green balls G_1, G_2, G_3. A trial of the experiment consists of simultaneously drawing two balls at random from the urn
- The outcomes of this experiment are subsets of size 2 of the form $\{R_1,R_5\}$ or $\{R_4,G_1\}$ or $\{G_2,G_3\}$

How many subsets of each kind?

- There are $\binom{9}{2}$ = 36 subsets of size 2 from a set of 9 balls, and the collection of these subsets is the sample space
- $\binom{6}{2}$ = 15 outcomes consist of two red balls
- $\binom{3}{2}$ = 3 outcomes have two green balls
- $\binom{6}{1}\binom{3}{1}$ =18 outcomes have 1 red, 1 green

More probabilities

- A = $\{R_2$ is in the sample drawn$\}$
- B = $\{G_2$ is in the sample drawn$\}$
- What is P(A B) ?
- P(A) = 8/36 P(B) = 8/36
- AB = {outcome = $\{R_2, G_2\}$} is a singleton event, and hence P(AB) = 1/36
- P(A B) = P(A) + P(B) – P(AB) = 15/36
- Exercise: What is $P(A^c$ $B^c)$?

Sampling without replacement

- A sample of size k from a set of size n is a subset of size k
- $\{R_2, G_2\}$ is the same as $\{G_2, R_2\}$
- Instead of obtaining the k elements of the subset simultaneously, we could draw them out one at a time, each draw being carried out without replacing the previously drawn elements
- This is sampling without replacement

Sampling without replacement

- If a previously drawn element were to be put back, then it could be drawn again
- This is not what we want!
- Sampling without replacement results in the same subsets of size k as random sampling, but the elements in the subset are arranged in a specific order
- Outcomes should be viewed as vectors: (R_2, G_2) is not the same as (G_2, R_2)

Sampling without replacement

- A subset of size k can be arranged into k! different vectors
- Outcomes of the experiment are $k!\binom{n}{k}$ $= n(n–1)(n–2)…(n–k+1)$ vectors of length k whose entries are k distinct elements from the set of size n
- Usual explanation: n choices for first entry, then n–1 for second, n–2 for third, etc.

Sampling without replacement

- Example: Our urn contains 6 identical red balls R_1, R_2, R_3, R_5, R_5, R_6 and 3 identical green balls G_1, G_2, G_3. Two balls are drawn at random without replacement from the urn
- 72 different samples without replacement can be drawn from the urn
- As usual, at random means that each of the 72 outcomes is equally likely

A surprising result?

- P{first ball is green}
  = P{outcome is (G_1, ❋)}
  + P{outcome is (G_2, ❋)}
  + P{outcome is (G_3, ❋)}
  = (8 + 8 + 8)/72 = 1/3
- P{second ball drawn is green} = P{(❋,G_1)} + P{(❋,G_2)} + P{(❋,G_3)} = 1/3 also!
- Here, ❋ and ❋ denote any ball other than the G_i that is exhibited

Surprising result is quite general

- More specifically, P{first ball drawn is G_2} = P{outcome is (G_2, ❋)} = 8/72 = 1/9
- P{second ball drawn is G_2} = P{(❋,G_2)} 8/72 = 1/9 also!
- More generally, if (x_1, x_2, … x_k) is the random sample without replacement, and A is some subset (e.g. “green ball” or {G_2}) of the set of n elements, then
  P{x_1 A} = P{x_2 A} = … = P{x_k A}

P{both balls are green}?

- P{both balls are green}
  = P{(G_1, G_2), (G_2, G_1), (G_1, G_3), (G_3, G_1), (G_2, G_3), (G_3, G_2)}
  = 6/72 = 3/36 as with a random sample
- Sampling without replacement is a “closer look” at a random sample: the ordering is important in the former, not in the latter
- When the ordering is irrelevant, as in {both balls are green}, we should get the same result, and we did

Problems with classical approach

- The classical approach of equally likely outcomes cannot be applied to sample spaces with infinitely many outcomes
- The nonclassical approach does extend to some, but not all types, of infinite sample spaces
- Can a sample space have infinitely many outcomes?
- Yes

Infinitely many outcomes

- Example: The experiment consists of tossing a coin till a Tail appears for the first time. The sample space is Ω = {T, HT, HHT, HHHT, … }
- This sample space is countably infinite
- Countable means there is a one-to-one correspondence between the integers and the outcomes

n–1 heads

Integer n ↔ HHH…HT

It's not in my job description!

- Automatic Repeat Request (ARQ) communication systems transmit a data packet (with error-detection mechanism such as a CRC checksum)
- The receiver replies with an NACK or ACK according as it detects errors or not
- Sender repeats transmission if the receiver replies NACK
- Do it over and over again till it's done right!

Countably infinite sample spaces

- Let $\Omega = \{x_1, x_2, \ldots x_n, \ldots\}$ be the countably infinite sample space
- $P\{x_n\} = p_n$ where $p_n \geq 0$
- For a finite subset A of Ω, P(A) is just the sum of the probabilities of the outcomes comprising the event A, as before
- It seems reasonable to have this idea work for an infinite subset of A as well
- But we need a new improved axiom

Axiom III needs improvement

- Axiom III for a finite sample space states that P(A B) = P(A) + P(B) for disjoint events A and B
- For a finite set of disjoint events, this readily extends to P(A B … G) = P(A) + P(B) + …+ P(G), but it does not extend (e.g. via mathematical induction) to an infinite collection of disjoint events
- Mathematical details are too technical

New improved Axiom III

- New improved Axiom III: Let $A_1, A_2, \ldots A_n, \ldots$ denote a countable sequence of disjoint events, that is, $A_i \cap A_j =$ for all $i \neq j$. Then, $P(A_1 \; A_2 \; \ldots \; A_n \; \ldots)$ $= P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots$
- The new Axiom III implies that P( ) = 0 and P(A B) = P(A) + P(B) for AB =
- See Ross, p. 31 for details

Better living through new Axiom III

- For a finite subset A of Ω, P(A) is just the sum of the probabilities of the outcomes comprising the event A, as before
- The new improved Axiom III extends this to infinite subsets as well
- With the new improved Axiom III, we can treat countably infinite sample spaces just as we handle finite sample spaces

Some thoughts about the p_n

- Special case: $\Omega = \{x_1, x_2, \ldots x_n, \ldots\}$
  $= \{x_1\} \quad \{x_2\} \quad \ldots \quad \{x_n\} \quad$ …and hence
  $P(\Omega) = P\{x_1\} + P\{x_2\} + \ldots + P\{x_n\} + \ldots$
  $= p_1 + p_2 + \ldots + p_n + \ldots = 1$
- The p_i's cannot all be equal (Why not?)
- The sequence of p_n's converges to 0
- The series $p_1 + p_2 + \ldots + p_n + \ldots$ converges to 1, and any subseries of this series also converges

An example

- $\Omega = \{x_1, x_2, \ldots x_n, \ldots\}$ where
  $p_n = P\{x_n\} = (0.6)\times(0.4)^{n-1}$ for all n
- $p_1 + p_2 + \ldots + p_n + \ldots$
  $= (0.6) \times [1 + (0.4) + \ldots + (0.4)^{n-1} + \ldots]$
  $= (0.6) \times 1/[1 - 0.4] = 1$ on summing the geometric series
- $1/[1-x] = 1 + x + x^2 + \ldots + x^n + \ldots$
  for $|x| < 1$ is one of the most useful results in this course. Memorize it!

Example (continued)

- $\Omega = \{x_1, x_2, \ldots x_n, \ldots\}$ where
  $p_n = P\{x_n\} = (0.6)\times(0.4)^{n-1}$ for all n
- Let $A = \{x_2, x_4, \ldots x_{2k}, \ldots\}$ denote the event that the outcome of the experiment is x_i where i is an even integer
- $P(A) = p_2 + p_4 + \ldots + p_{2k} + \ldots$
  $= (0.6)\times(0.4)\left[1 + (0.4)^2 + (0.4)^4 + \ldots\right]$
  $= (0.24)\times 1/[1 - (0.4)^2] = 2/7$

What do probabilities mean?

- The classical approach to probability justifies its equally likely outcomes with an appeal to symmetry
- The nonclassical approach does not necessarily provide reasons for its assignments of probabilities
- So, why should different outcomes have different probabilities? What is the reason for setting $P(x) = 0.1$ and $P(y) = 0.2$?

Observed relative frequencies

- Probabilities are expressions of belief about what is going to happen in the future
- One reason for setting $P(x) = 0.1$ and $P(y) = 0.2$ might be based on what was observed in the past
- If outcome x was observed N_x times on N trials of the experiment, then N_x/N is called the observed relative frequency of x
- For events: N_A = # of occurrences of A

Relative frequency = probability?

- Observed relative frequencies satisfy the probability axioms
- Event A was observed N_A times on N trials where $N_A \geq 0$. Hence, $N_A/N \geq 0$
- Ω was observed on all trials: $N_\Omega/N = 1$
- If A and B are disjoint events, then on the trials on which A occurred, B could not have occurred, and vice versa. Hence,
  $N_{A \quad B} = N_A + N_B \qquad N_{A \quad B}/N = N_A/N + N_B/N$

Estimates of probabilities

- Based on past experimental results, we can use the observed relative frequency as an estimate of the probability of an event
- Estimates are usually reasonably good if the number of trials was large
- On N trials, the relative frequency that we observe might easily differ from the true probability by as much as $N^{-1/2}$ or more

But they don't always work!

- For a sample space with countably infinite outcomes, at most N different outcomes will be observed on N trials
- The probabilities to be assigned to the infinitely many unobserved outcomes is again a matter of guesswork!
- Since the sequence of p_n's converges to 0, most of the probabilities are much smaller than $N^{-1/2.}$ Estimation errors?

A million tosses, a million laughs

- Suppose that a fair coin is tossed a million times
- Is there a logical reason why the coin will not turn up Heads each and every time?
- No, there is no logical reason why it couldn't, but it is very unlikely to do so
- Yes, if the coin is fair, there is no way that it can turn up Heads a million times in a row

Faith versus belief

- A fair coin that is tossed a million times will not turn up Heads a million times in a row?
- What is the largest number of consecutive Heads that you think can possibly occur?
- How does the coin know that you are watching, and therefore, it should turn up Tails after a certain number of Heads?
- What if more than one person is watching?

No guarantees

- A million Heads in a row might later be followed by a million Tails to give a relative frequency of 1/2 for Heads
- If only we had gone one for another million tosses, we would have been OK …
- No guarantee that an observed relative frequency is close to the actual probability
- We expect the relative frequency is close, but we cannot guarantee it

Estimates only

- Observed relative frequencies can serve only as estimates of probabilities
- Do not use them to define probabilities even though they satisfy the axioms
- No guarantee that an observed relative frequency is close to the actual probability, only a strong expectation
- "What does probability mean?" It is a numerical expression of strength of belief

Large number of trials

- In a large number of trials (say coin tosses) long runs of Heads (or Tails) can be expected to occur
- Actual coin tosses vs human simulation
- The gambler's ruin: If #Heads > #Tails (say) at any particular point in the game, then this lead will last for a long time
- Lead changes become more infrequent
- Deviations increase in absolute value

Limit of relative frequency?

- A lot of effort was expended in trying to define probability as the limit of the relative frequency
- $P(A) = \lim_{N \to \infty} N_A/N$
- Unfortunately, the limit does not exist in a mathematical sense
- Physically we will only observe a finite-length prologue of the sequence of trials

Uncountably infinite sample spaces

- The nonclassical approach cannot be used at all for uncountably infinite sample spaces
- Example: Pick a random number between 0 and 1
- $\Omega = \{x : 0 < x < 1\}$
- Intuitively, the meaning of random in this instance is that we do not favor any one number in the interval (0,1) over the others

Example (continued)

- Example: Pick a random number between 0 and 1
- $\Omega = \{x : 0 < x < 1\} = (0, 1)$
- One way of expressing the innate randomness of the choice is as follows:
- Given any subinterval of (0, 1), the probability that the chosen number lies in that subinterval is equal to the length of the interval

God made the integers

- Kronecker: God made the integers; all else is the work of man
- Human beings usually choose rational numbers when asked for a number in (0,1)
- A physical measurement made with an instrument will yield a rational number
- `rand()` returns "real numbers" that are actually rational numbers
- All this is because of finite precision

Do real numbers exist?

- The real number line is a mathematical construction that models the real world very well indeed
- If the volume electrical charge density is ρ, the charge in a volume Δv is just $\rho\Delta v$
- For Δv very small, $\rho\Delta v$ is smaller than the charge of an electron, so the model cannot be right for small volumes (or densities)!
- But it is convenient!

Physical Science and Engineering

- In the physical sciences and engineering, the real numbers are a model for very many phenomena that are discrete at the microscopic level
- This usually causes no problems and the model usually gives the correct answers
- Calculus can be applied
- We all understand that V = 1.235 volts really means 1.2345 ≤ V ≤ 1.2355 volts

Real numbers in probability

- The uncountably infinite sample spaces that are intervals of the real line (or the entire real line) cause many subtle mathematical difficulties in probability theory
- There are some obvious problems as well
- The resolution of these difficulties led to the development of the axiomatic theory of probability

What is P{outcome}?

- The output of `rand()` is a good model for repeated trials for the experiment of picking a number at random in (0, 1)
- Consider a million random numbers obained from `rand()`. A particular outcome, say 0.703546789, will either not have occurred in these million trials, or it will have occurred just once.
- P{0.703546789} ≈ 10^{-6} or 0

P{outcome} is always 0?

- The relative frequency estimate of P{0.703546789} ≈ 10^{-6} or 0
- If `rand()` is invoked a further million times, then 0.703546789 will have occurred a total of 0 or 1 times in 2 million trials P{0.703546789} ≈ 0.5×10^{-6} or 0
- The relative frequency estimate seems to be converging to 0 as the number of trials increases!

P{outcome} is always 0

- The only model that works for uncountably infinite sample spaces is for each outcome to have probability 0
- But, on each trial, some outcome occurs, doesn't it?
- So where are the probabilities?
- For `rand()`, P{a < outcome < b} = b–a
- The nonzero probabilities are assigned to the intervals of the line, not to outcomes!

Summary

- We studied sampling without replacement and its relation to random samples
- We studied sample spaces with countably infinite outcomes and adopted a new Axiom III
- We waxed philosophical about relative frequencies, estimates, faiths and beliefs
- We began to study uncountably infinite sample spaces and noted P{outcome}=0