Monte Carlo Methods – Background

What is a Monte-Carlo method?

- In a Monte-Carlo method, the desired answer is formulated as a quantity in a stochastic model
- Model is fed inputs drawn from sampling a distribution randomly
- Example: Cube Dies, with the sides numbered 1 to 6. Determine if “fair”
  - toss the cube 120 times,
  - observe which side comes up on top each time
  - study whether the sides occur with approximately equal frequency
  - How do we decide if our experiment answered the question?

- Example: “black box” that takes input a number between 0 and 1 as input and emits a number between 0 and 1 as an output
- Characterize the “process” that relates input output relations
  - Feed the box m numbers and observe the m outputs of the box
  - What measures to use?
  - average of the observations as an estimate of the statistical mean of the process defined by the black box for each input
  - Here we are estimating the outputs for each input, so estimating a function

Why Monte Carlo Methods?

- Ideas of sampling and distributions important in many applications
  - E.g., testing on representative inputs
  - Simulation to get outputs
  - Statistical Physics
  - Economics
  - Optimization
- Area of mathematics used in very complex systems with relatively simple outputs
  - Networking (computer and social)
  - Modeling humans or artificial systems
  - Thermodynamics
- Leads on to more powerful analysis tools and modeling paradigms
  - Markov chains
  - Metropolis Hastings

Terminology

- **Monte Carlo methods:** estimate quantities by random sampling
- **pseudo-Monte Carlo methods:** Use samples that are more systematically chosen
- All practical computational methods are pseudo-Monte Carlo, since random number generators implemented on machines are generally not truly random.
- We’ll use the term Monte Carlo for samples that are generated using pseudorandom numbers generated by a computer

Probability

- Observe some set process and events
- Formalization of “likelihood” of an event
- Event space
- “Equally likely” events
- Ratio of event counts to total
  - Zero indicates impossibility
  - One indicates certainty
- Total probability
  - *Some* outcome must occur, so total of all probabilities must be 1
- Ultimately, an element of “faith” or “trust” in nature

Philosophy

- Engineers/Mathematicians: Often view Monte Carlo methods as methods of last resort
- They are generally quite expensive and only applied to problems that are too difficult to handle by deterministic (non-stochastic) methods
- Machine learning, statistical physics and Bayesian statisticians view it as a fundamental way of viewing the world …
  - All scientific laws are derived from observation

Chapters in the Book

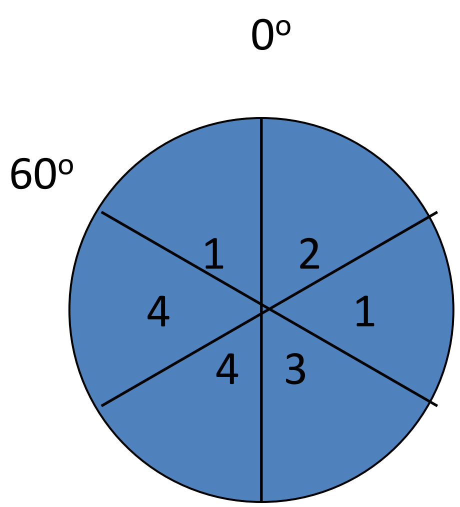
- Basic statistics: Random and pseudorandom numbers and their generation: Chapter 16.
- Monte Carlo methods for numerical integration: Chapter 18.
- Monte Carlo methods for optimization: Chapter 17.
- An example of Monte Carlo methods for counting: Chapter 17.

Basic statistics

- Basic statistics: Random and pseudorandom numbers and their generation
- What is a random number?
- What are the mean and variance of a random sample?
- What is a distribution? What are its mean and variance?
- How are pseudorandom numbers generated?

Random Variable (RV)

- A mapping from events to a number
  - Probability of an *event* occurring translates to
  - Probability of a RV having a particular *value*
  - Probability of a particular *outcome* to an *experiment* or a *trial*
- With the wheel shown,
  - Each sector is “equally likely”
  - “Pointer stops between 60 and 180” is equivalent to
  - “X receives the value 4”



Some random number generators

- **Raffle**: Take n papers and number them 1 to n. Put them in a box, and draw one at random. After you record the resulting number, put the paper back in the box. You are taking random numbers that are uniformly distributed among the values {1, 2, . . . , n}.
- **Roulette Wheel or Spinner**: Make a spinner by anchoring a needle at the center of a circle. Draw a radius line on the circle. Spin the needle, and measure the angle it forms with the radius line. You obtain random numbers that are uniformly distributed on the interval $[0, 2\pi)$.
- **Printed Tables:** There are printed tables of random numbers.
- **Natural Process:** A radioactive substance emits -particles every μ seconds, then the time between two successive emissions has the exponential distribution with mean μ. (Note: This is a special case of the Gamma distribution.)
- “Distribution”?

Introductory Concepts

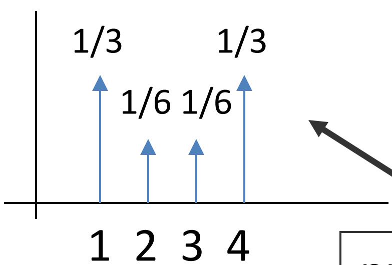
- Distribution
- Density
- Expectation
- Conditional Probability
- Independent Events
- Memoryless RVs

Probability Functions

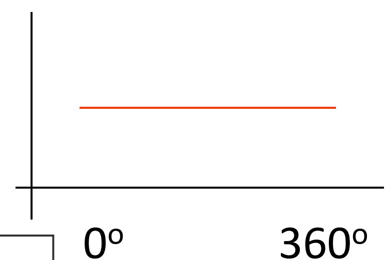
- Functions mapping value of an RV (outcome of a trial) to the probability of it occurring
  - Discrete (finite or infinite) set of outcomes described by probability mass function
  - Continuous set of outcomes described by probability density function
  - Sometime both are called density function
- Cumulative Density/Mass Function
  - Adding up probability over successive outcomes
  - Sometimes called Probability Distribution Function
- *Range* of RV is the *domain* of P [.]

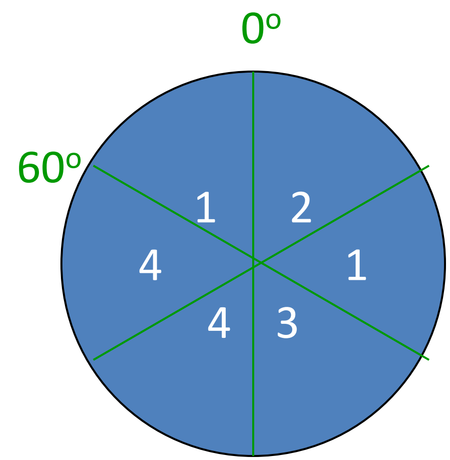
Roulette Wheel

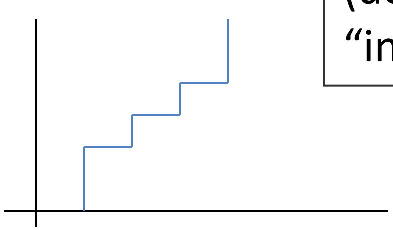
- The "place where pointer stops" is a *continuous* variable, say θ
- The RV X defined as "1 if θ is between 0 and 60, 4 if θ is between 60 and 180, …" is *discrete*
  - Mapping of event to number, the definition of RV
  - Range of X is {1, 2, 3, 4}
- Probability functions: mappings from range of X (or θ)

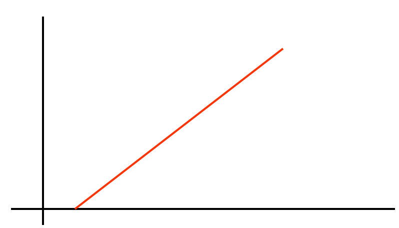


pmf is drawn with impulses (delta functions), to indicate "infinite" density





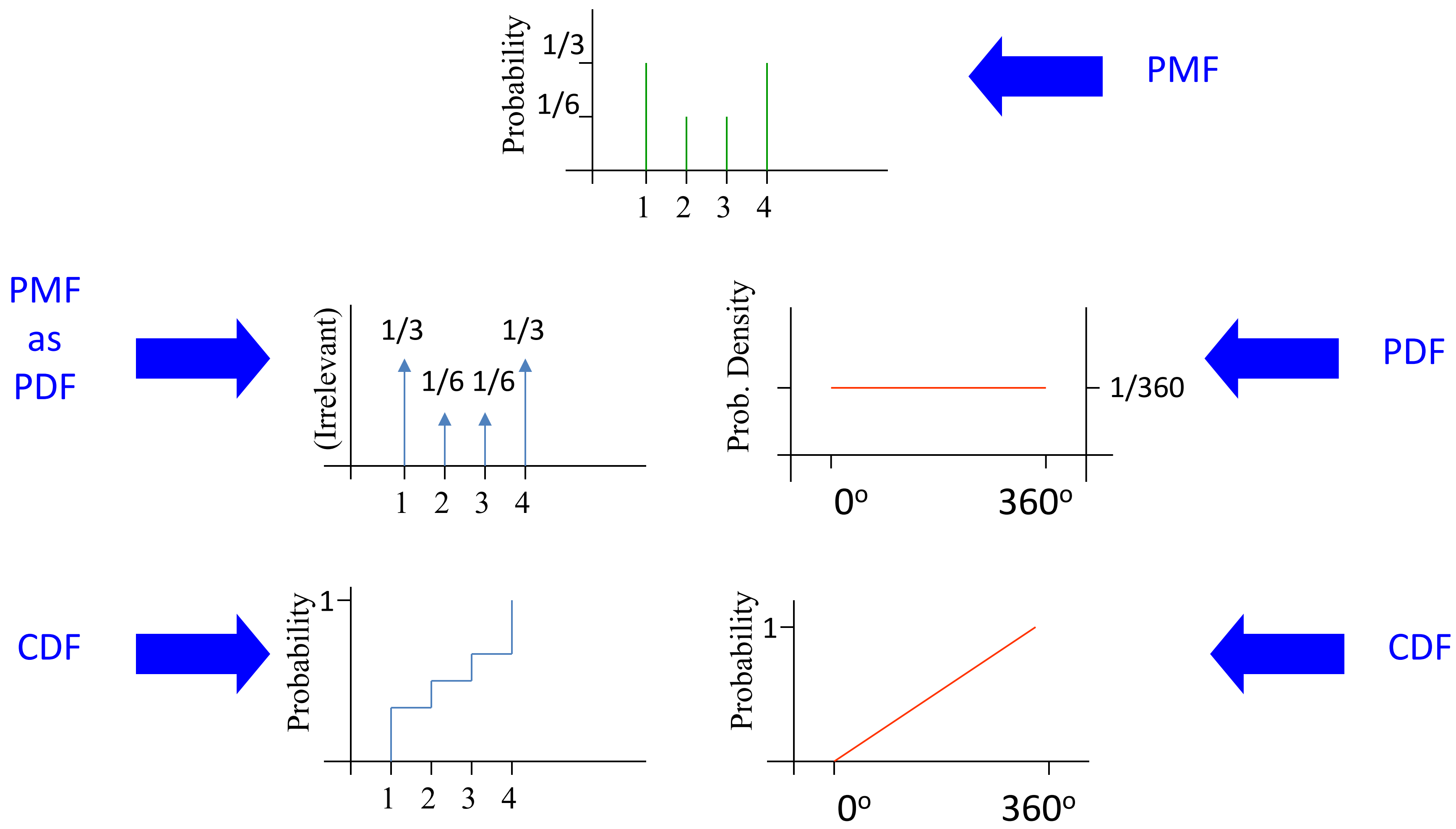




PMF, PDF, PMF as PDF, CDF

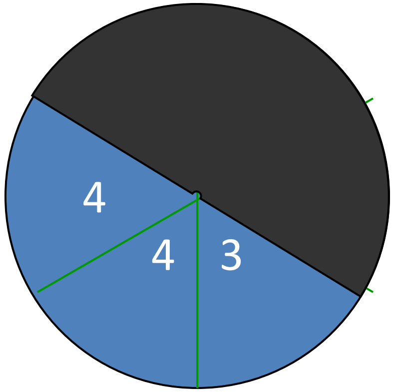
- PMF indicates actual probability at specific points
  - Probability is a pure (dimensionless) number (ratio of event counts)
- PDF indicates rate at which probability is accumulated (hence: density)
  - Derivative of probability (also a pure number!)
- Amassed probability (PMF) can be indicated on a PDF
  - But must resort to “infinite” densities (impulses)
- CDF is integral of PDF, so indicates actual (cumulative) probabilities at each point)

PMF, PDF, PMF as PDF, CDF (2)



Conditional Probability and Independence

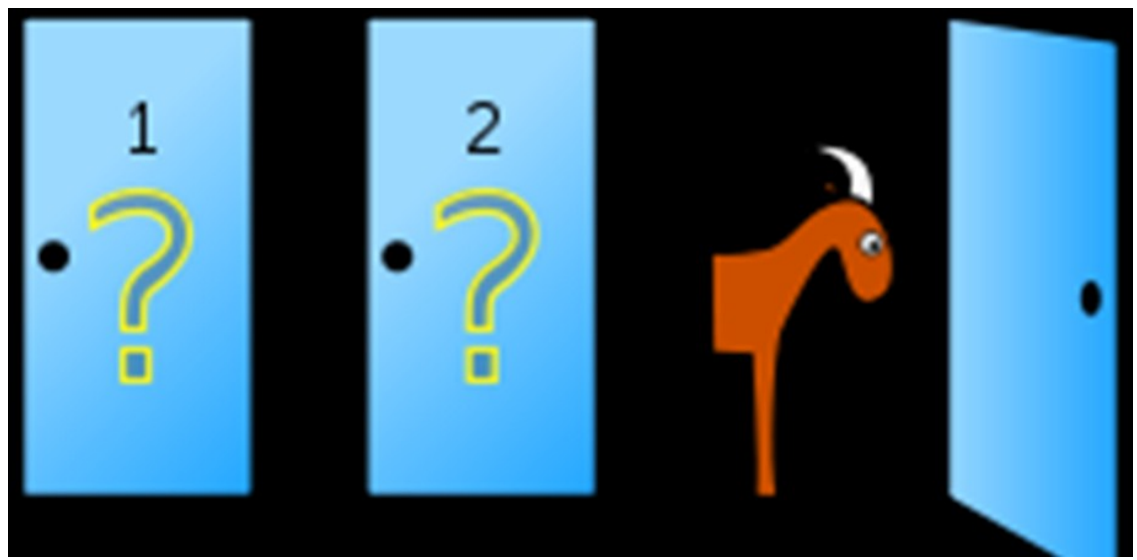
- Conditional Probability
  - Partial knowledge, "Given that…"
  - "Sub-universe"
  - P [A | B] = P [AB] / P [B]
    - A = "Get 4", B = "shaded area"



- Concept of *independence*
  - Partial knowledge is not relevant
  - Conditional and unconditional probabilities same
  - Also implies that P [A] P [B] = P[AB]
  - Which implies that E {AB} = E {A} E {B}

Conditional Probabilities - Tricky

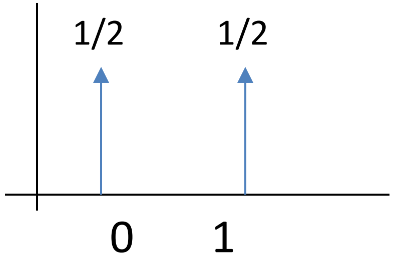
- "Monty Hall problem"
- Game show - three doors, one car and two goats
  - "Pick a door" - contestant picks one
  - Host opens one of the other two to reveal a goat
  - "Will you switch?"
- Should you?



Incredible but true - you should
Based on repeatable assumptions
Host always offers choice (or independently of your choice)
Host never opens the "car" door

The Bernoulli Distribution

- Basically, the "coin toss" distribution
- Two possible outcomes - one has a probability of p, the other $1 - p$
  - With fair coin, $p = 0.5$



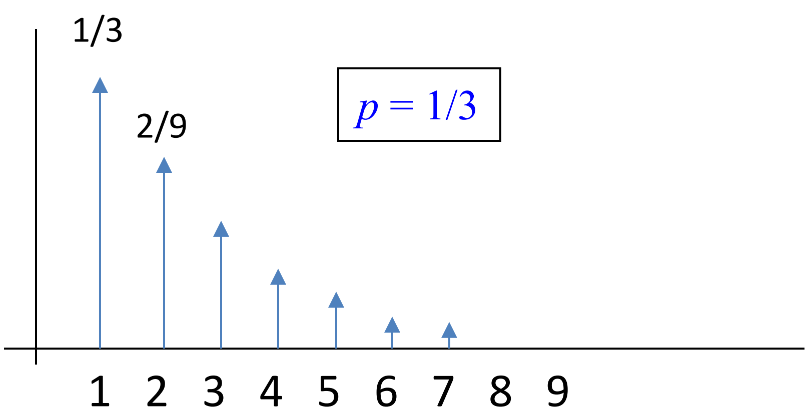
The Geometric Distribution

- “Try until a specific outcome” experiment on a event with Bernoulli distribution
- For example, consider the coin toss
  - Keep tossing until you get Heads
  - Value of RV = number of tosses it took
  - One trial of this RV involves multiple trials of Bernoulli RV
- What is the probability distribution of this RV?
  - Key concept - each coin toss is independent

Obtaining Geometric Distribution

- Probability that RV X will have value 1
  - Same as probability that first Bernoulli RV will be Heads = *p*
- Probability that X will have value 2
  - Compound event
  - Probability that first Bernoulli RV will NOT be Heads, AND second Bernoulli RV will be Heads
  - Since tosses are independent, can multiply probabilities

x	P [X = x]
1	*p*
2	
3	
…	
k	
…	



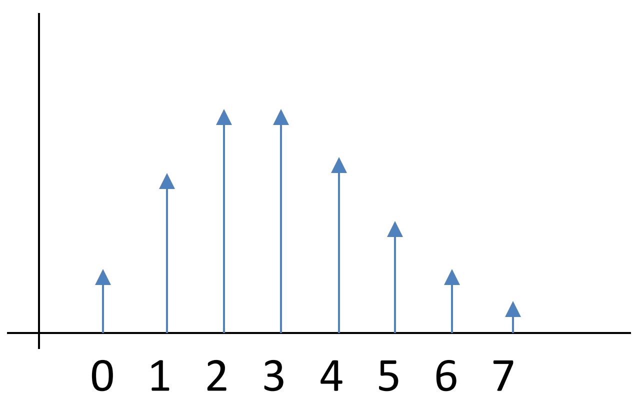
Next term? (2/3*2/3)*1/3=4/27

The Binomial Distribution

- The "*k* out of *n*" distribution
- Flip *n* coins, RV is the number of Heads
  - It does not matter whether coins are tossed simultaneously or in sequence (or one coin *n* times)
- Probability of the RV being 0
  - Compound event: Bernoulli RV 1 is Tails, *and* Bernoulli RV 2 is Tails, …
  - Easily, $(1-p)^n$
- Probability of RV being 1
  - Compounding of compound events
  - BRV1 is Heads, *and* BRV2 is Tails, *and* …
  - *OR*, BRV1 is Tails, *and* BRV2 is Heads, *and* …
  - *OR* …
- Probability of RV being 2
  - BRV1 is Heads, *and* BRV2 is Heads, *and* BRV3 is Tails, *and* …
  - *OR,* BRV1 is Heads, *and* BRV2 is Tails, *and* BRV3 is Heads, …
  - *OR*, …

The Binomial Distribution

- Probability of X = k
  - Probability of k BRVs being Heads *and* (n-k) BRVs being Tails
  - *Times* number of distinct ways in which you can pick k things out of n
- $P[X = k] = {}^{n}C_{k}\ p^{k} (1 - p)^{n-k}$



Normal or Gaussian Distribution

- A good model in many situations:
  - The pattern of leaves that fall from a symmetric tree
  - The IQ measure of intelligence was constructed so that the measures are normally distributed

  http://www.awl.com/weiss/e_iprojects/c06/chap06.htm
  - Physical characteristics of plants/animals (height, weight, etc.).
  - velocity distribution of molecules in thermodynamic equilibrium (Maxwell-Boltzmann distribution)

  http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/disfcn.html
  - Measures of psychological variables such as reading ability, introversion, job satisfaction, and memory.

  http://davidmlane.com/hyperstat/normal_distribution.html
- First studied by DeMoivre (1667-1754), in connection with predicting the probabilities in gambling.