ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

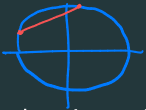
Unit 1: Probability Review

Sid Banerjee

School of ORIE, Cornell University

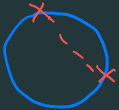
clicker question: random chords

Given the circle $x^2 + y^2 = 1$, we want to sample a uniform random chord. How can we do this?



(a) Pick 2 endpoints u.a.r on the circumference of the circle

8%



(b) Pick circumference $C(\theta)$ at uniform random angle $\theta \in [0, 2\pi]$, and pick chord-center u.a.r. on $C(\theta)$

10%

(c) Pick the chord-center u.a.r. inside the circle (draw radius to point, and chord perpendicular)

8%

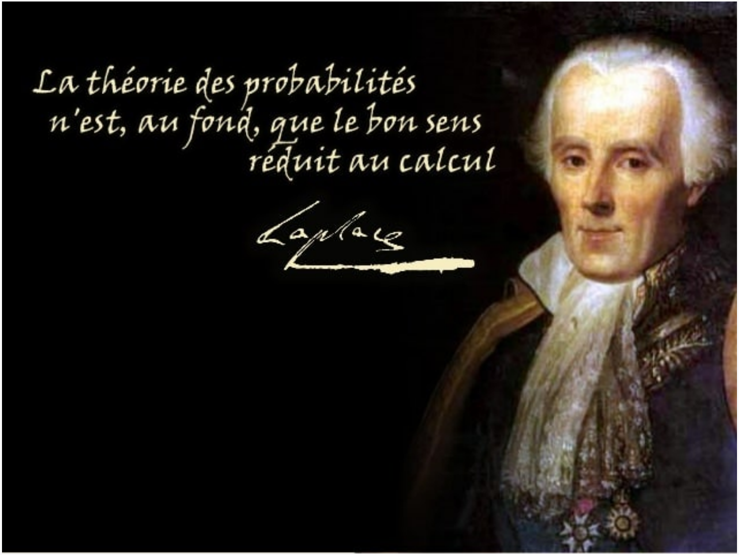
(b)

(c)

(d) All of these are the same

74%

Given the circle $x^2 + y^2 = 1$, we want to sample a uniform random chord. How can we do this?

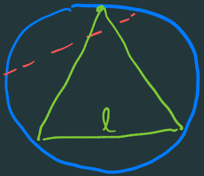


“probability theory is common sense reduced to calculation”

not quite...

Bertrand's problem

given equilateral triangle inscribed in a circle and a random chord, what is the $\mathbb{P}$[chord is longer than side of the triangle]?

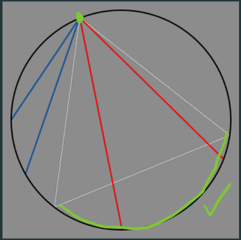


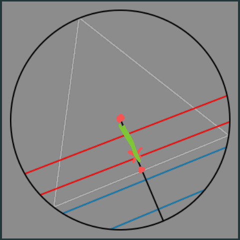
$\ell = \sqrt{3}\, r$

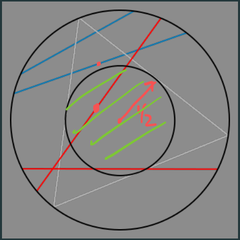
not quite...

Bertrand's problem

given equilateral triangle inscribed in a circle and a random chord, what is the $\mathbb{P}$[chord is longer than side of the triangle]?







$1/3$

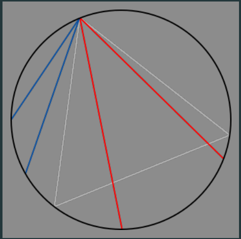
$1/2$

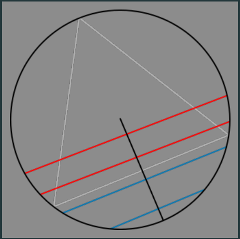
$$\frac{\pi r^2}{\pi} = \frac{1}{4}$$

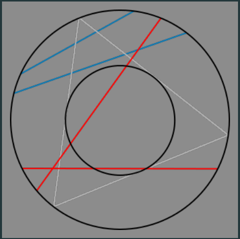
not quite...

Bertrand's problem

given equilateral triangle inscribed in a circle and a random chord, what is the $\mathbb{P}$[chord is longer than side of the triangle]?







the moral (for this course... and for life)

be very precise about defining experiments/random variables/distributions

also see Wikipedia article on Bertrand's paradox

the essentials

reading assignment

Bishop: chapter 1, sections 1.2 - 1.2.4

Mackay: chapter 2 (less formal, but much more fun!)

things you must know and understand

- random variables (rv) and cumulative distribution functions (cdf)
- conditional probabilities and Bayes rule
- expectation and variance of random variables
- independent and mutually exclusive events
- basic inequalities: union bound, Jensen, Markov/Chebyshev
- common rvs (Bernoulli, Binomial, Geometric, Gaussian (Normal))

the essentials

These are the things you must know and understand

- Random variables (rv) + Cumulative Distribution Fn (cdf)
- Expectation and Variance of random variables
- Independence (and also dependence – mutually exclusive events, conditioning, Bayes rule)
- Common rvs (Bernoulli, Binomial, Geometric, Gaussian, Exponential, Poisson)

random variables

random variables

sample space Ω: set of all possible outcomes of random expmt

random variable: any function from $\Omega \to \mathbb{R}$

Outcome is unknown
beforehand

random variables

sample space Ω: set of all possible outcomes of random expmt

random variable: any function from $\Omega \to \mathbb{R}$

example: Youtube's ad algo (I think!) – pick a random number of ads between 0 and 2 (inclusive), and a random length between 0 and 30s for each ad. Let T = Total length of ads on video

$$\Omega = \underbrace{\{0,1,2\}}_{n} \times \overset{t_1}{[0,30]} \times \overset{t_2}{[0,30]}$$

discrete

Cartesian product

continuous

$$T(n, t_1, t_2) = \sum_{i=0}^{n} t_i$$

probabilities

\mathbb{P}

Probability $\mathbb{P}(E)$: 'number' for 'every' subset $E \subseteq \Omega$, such that: $\mathbb{P}(E)$

- $\mathbb{P}(E) \geq 0$ for all $E \subset \Omega$ non-negative
- $\mathbb{P}(\Omega) = 1$ (i.e., probs 'summed over all outcomes' adds to 1)
- $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$ *
  (i.e., probs add for mutually exclusive events $E_1 \cap E_2 = \varphi$)

mutually exclusive
("if E_1 happens, E_2 can't happen")

cumulative distribution function

ALERT!!

always try to think of probability and rvs through the cdf

- for any rv X (discrete or continuous), its probability distribution is defined by its cumulative distribution function (cdf)

$$F(x) = \mathbb{P}[\underbrace{X \leq x}_{E}] \qquad \mathbb{P}[\{\omega \in \Omega \mid X(\omega) \leq x\}]$$

- using the cdf we can compute probabilities

$$\mathbb{P}[a < X \leq b] = F(b) - F(a)$$

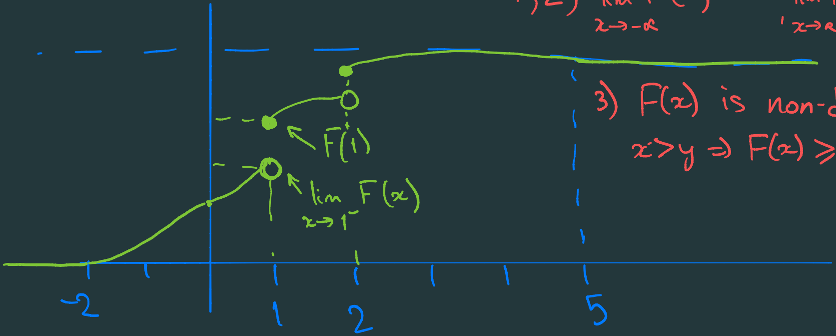
$X \in (a, b]$

visualizing a cdf

The plot of a cdf obeys 3 essential rules + one convention

Example: consider an rv $\in [-2, 5]$ with a **jumps** at 1 and 2

1,2) $\lim_{x \to -\infty} F(x) = 0$ $\lim_{x \to \infty} F(x) = 1$



3) $F(x)$ is non-decreasing

$x > y \Rightarrow F(x) \geq F(y)$

Convention – 'Closed circle from the right'

(because $F(x) = \mathbb{P}[X \leq x]$)

discrete rv

- for a discrete random variable, another characterization is its probability mass function (pmf) $p(\cdot)$ $X \in \{1, 2, 3, \dots\}$

$$p(x) = \mathbb{P}[X = x]$$

- The pmf $p(\cdot)$ is related to the cdf $F(\cdot)$ as

$$F(x) = \sum_{y \leq x} p(y)$$

$$p(x) = F(x) - F(x-1)$$

largest value smaller than x that X can take

- further, any pmf $p(x)$ obeys 2 properties:

$$p(x) \geq 0 \;\forall x, \quad \sum_{x=1}^{\infty} p(x) = 1$$

continuous random variables

- for a continuous random variable taking values in $\mathbb{R}$, another characterization is its probability density function (pdf) $f(\cdot)$

$$\mathbb{P}[a < X \leq b] = \int_a^b f(x)\,dx = F(b) - F(a)$$

$$X \in (a,b]$$

- any pdf $f(x)$ obeys 2 properties:

$$f(x) \geq 0 \ \forall x, \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

- ALERT!! not true that $f(x) = \mathbb{P}[X = x]$. In fact, for any x,

$$\mathbb{P}[X = x] = 0$$

continuous random variables

- for continuous rv X with pdf $f(\cdot)$ and cdf $F(\cdot)$, we have

$$\mathbb{P}[a < X \leq b] = F(b) - F(a) = \int_a^b f(x)dx$$

- now we can go from one function to the other as

$$F(x) = \int_{-\infty}^{x} f(x)\,dx$$

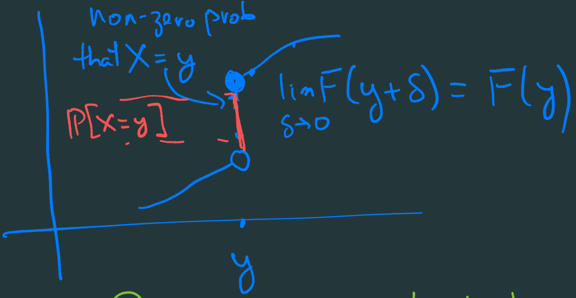
$$f(x) = \frac{d}{dx} F(x) \quad \text{(If } F \text{ is differentiable)}$$

note on end-points

we wrote: $\mathbb{P}[a < X \leq b] = F(b) - F(a)$: is $<$ vs $\leq$ important?

Convention - $F(x) = \mathbb{P}[X \leq x]$

no if F is cont
yes if discont

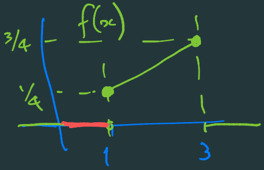


RCLL - Right Continuous Left Limit

clicker question: practice with cdfs

The probability density function of the rv X is given by:

$$f(x) = \begin{cases} x/4 & \text{if } 1 \leq x \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$



(a) $F(x) = \frac{1}{4}$ for $x \geq 0$

(b) $F(x) = \frac{x^2}{8}$ for $x \geq 0$ 45%

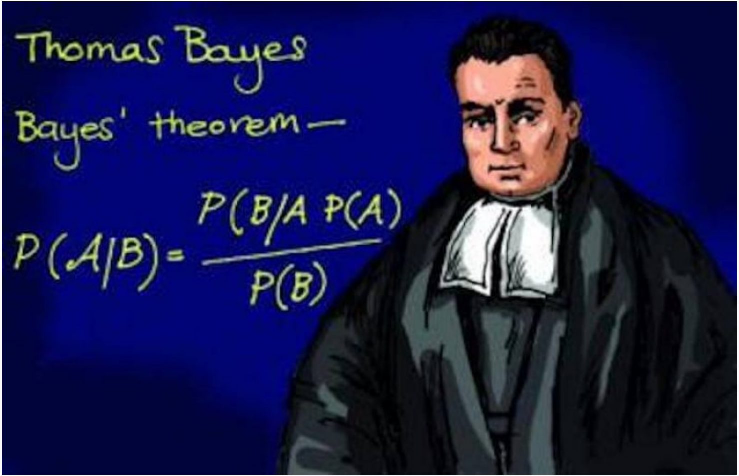
(c) $F(x) = \frac{x^2-1}{8}$ for $x \geq 0$ 30%

(d) None of these 20%

clicker question: solution

The probability density function of the rv X is given by:

$$f(x) = \begin{cases} x/4 & \text{if } 1 \leq x \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$



conditioning and Bayes' rule

marginals and conditionals

let X, Y be discrete rvs taking values in $\mathbb{N}$. denote the joint pmf:

$$p_{XY}(x, y) = \mathbb{P}[X = x, Y = y]$$

marginalization: computing individual pmfs from joint pmfs as

$$p_X(x) = \sum_{y \in \mathbb{N}} p_{XY}(x, y) \qquad p_Y(y) = \sum_{x \in \mathbb{N}} p_{XY}(x, y)$$

conditioning: pmf of X given $Y = y$ (with $p_Y(y) > 0$) defined as:

$$\mathbb{P}[X = x | Y = y] \triangleq p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

more generally, can define $\mathbb{P}[X \in \mathcal{A} | Y \in \mathcal{B}]$ for sets $\mathcal{A}, \mathcal{B} \in \mathbb{N}$

see also this visual demonstration

Bayesian inference

let X, Y be discrete rvs taking values in $\mathbb{N}$, with joint pmf $p(x, y)$

product rule

for $x, y \in \mathbb{N}$, we have: $p_{XY}(x, y) = p_X(x) p_{Y|X}(y|x) = p_Y(y) p_{X|Y}(x|y)$

sum rule

for $x \in \mathbb{N}$, we have: $p_X(x) = \sum_{y \in \mathbb{N}} p_{X|Y}(x|y) p_Y(y)$

Bayes rule

for any $x, y \in \mathbb{N}$, we have:

$$p_{X|Y}(x|y) = \frac{p_X(x) p_{Y|X}(y|x)}{\sum_{x \in \mathbb{N}} p_{Y|X}(y|x) p_X(x)}$$

see also this video for an intuitive take on Bayes rule

clicker question: Bayesian inference

Mackay's three cards

We have 3 cards $C1, C2, C3$, with $C1$ having faces Red-Green, $C2$ having faces Green-Green; and $C3$ having faces Red-Red.

A card is randomly drawn and placed on a table – its upper face is Red. What is $\mathbb{P}$[lower face is Red]?

(a) 0

Monty Hall

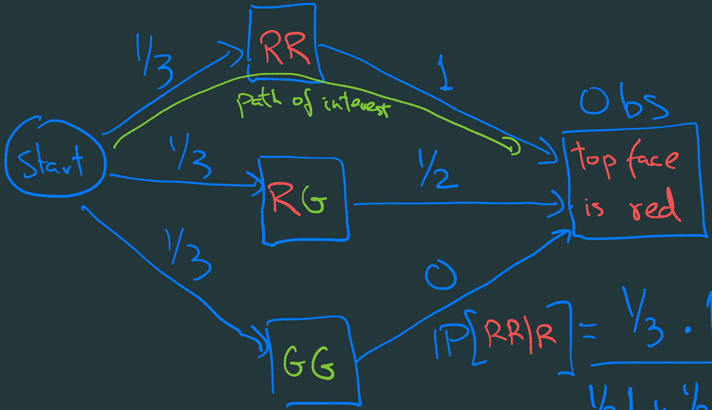
(b) 1/3 15%

(c) 1/2 70%

(d) 2/3 12%

(e) None of these

clicker question: Bayesian inference



$$\mathbb{P}[RR \mid R] = \frac{1/3 \cdot 1}{1/3 \cdot 1 + 1/3 \cdot 1/2 + 1/3 \cdot 0}$$

mutually exclusive

$$= \frac{2}{3}$$

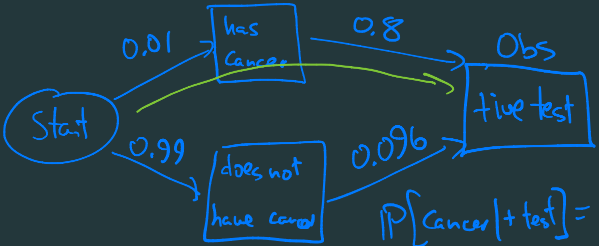
Bayesian inference: example

Eddy's mammogram problem

– $\mathbb{P}[\text{women at age 40 have breast cancer}] = 0.01$ true positive

– A mammogram detects the disease 80% of the time, but also mis-classifies healthy patients 9.6% of the time. false positive

If a 40-year old woman has a positive mammogram test, what is the probability she has breast cancer?



$$\mathbb{P}[\text{cancer} \mid +\text{test}] = \frac{8 \times 10^{-3}}{8\times10^{-3} + \underbrace{0.99\times 0.096}_{\gg 8\times10^{-3}}}$$

Bayesian inference: example

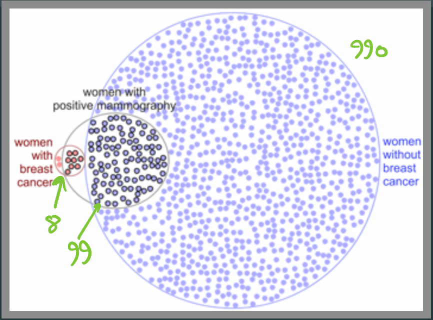
Eddy's mammogram problem

– $\mathbb{P}$[women at age 40 have breast cancer] $= 0.01$

– A mammogram detects the disease 80% of the time, but also mis-classifies healthy patients 9.6% of the time.

If a 40-year old woman has a positive mammogram test, what is the probability she has breast cancer?

Always compute the prob of everything!



credit: Micallef et al.

expectations and independence

expected value (mean, average)

let X be a random variable, and $g(\cdot)$ be any real-valued function

- If X is a discrete rv with $\Omega = \mathbb{Z}$ and pmf $p(\cdot)$, then

$$\mathbb{E}[X] = \sum_x x\, p(x)$$

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

'Law of large numbers'

- If X is a continuous rv with $\Omega = \mathbb{R}$ and pdf $f(\cdot)$, then

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$