

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

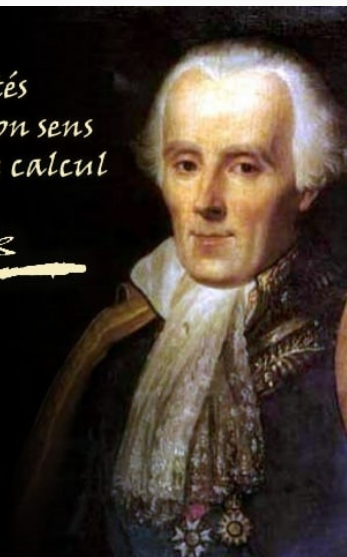
Given the circle $x^2 + y^2 = 1$, we want to sample a **uniform random chord**. How can we do this?

- (a) Pick 2 endpoints u.a.r on the circumference of the circle
- (b) Pick circumference $C(\theta)$ at uniform random angle $\theta \in [0, 2\pi]$, and pick chord-center u.a.r. on $C(\theta)$
- (c) Pick the chord-center u.a.r. inside the circle (draw radius to point, and chord perpendicular)
- (d) All of these are the same

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

*La théorie des probabilités
n'est, au fond, que le bon sens
réduit au calcul*

Laplace



“probability theory is common sense reduced to calculation”

not quite...

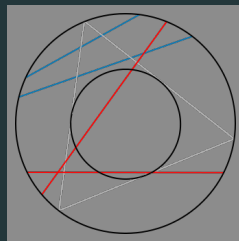
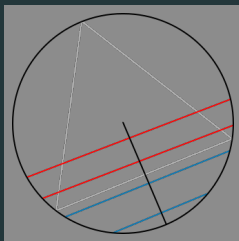
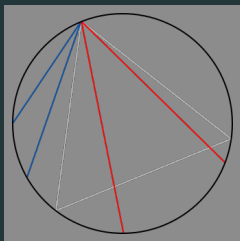
Bertrand's problem

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

not quite...

Bertrand's problem

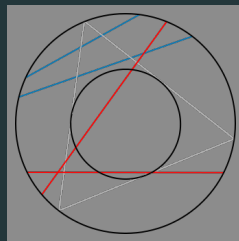
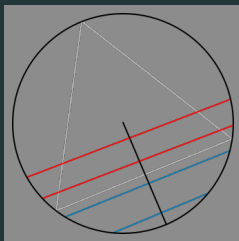
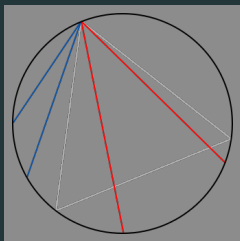
given equilateral triangle inscribed in a circle and a **random chord**, what is the $\mathbb{P}[\text{chord is longer than side of the triangle}]$?



not quite...

Bertrand's problem

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



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

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 $\Omega \rightarrow \mathbb{R}$

random variables

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

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

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 $T = \text{Total length of ads on video}$

probabilities

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

- $\mathbb{P}(E) \geq 0$ for all $E \subset \Omega$
- $\mathbb{P}(\Omega) = 1$ (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 $E_1 \cap E_2 = \varphi$)

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)**

$$F(x) =$$

- using the cdf we can compute probabilities

$$\mathbb{P}[a < X \leq b] =$$

visualizing a cdf

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

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

discrete rv

- for a discrete random variable, another characterization is its probability mass function (pmf) $p(\cdot)$

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

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

$$F(x) =$$

$$p(x) =$$

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

continuous random variables

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

$$\mathbb{P}[a < X \leq b] =$$

- any pdf $f(x)$ obeys 2 properties:
- **ALERT!!** not true that $f(x) = \mathbb{P}[X = x]$. In fact, for any x ,

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

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

$$F(x) =$$

$$f(x) =$$

note on end-points

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

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 $x \geq 0$

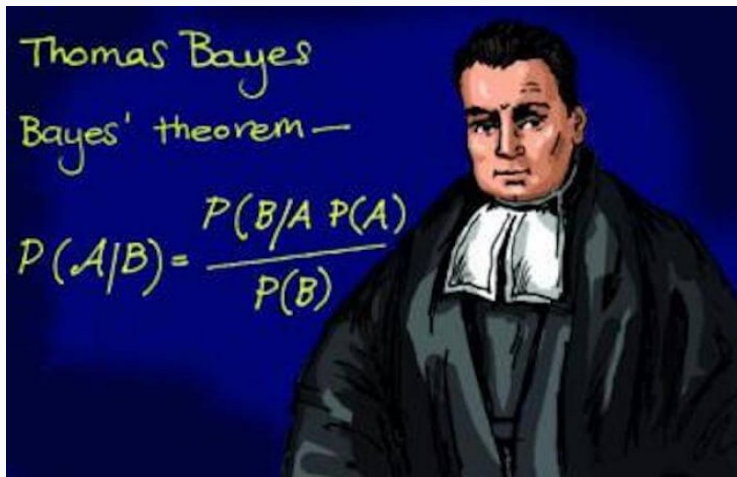
(c) $F(x) = \frac{x^2-1}{8}$ for $x \geq 0$

(d) None of these

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

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

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 C_1, C_2, C_3 , with C_1 having faces Red-Green, C_2 having faces Green-Green; and C_3 having faces Red-Red.

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

- (a) 0
- (b) $1/3$
- (c) $1/2$
- (d) $2/3$
- (e) None of these

clicker question: Bayesian inference

Bayesian inference: example

Eddy's mammogram problem

- $\mathbb{P}[\text{women at age 40 have breast cancer}] = 0.01$
- A mammogram detects the disease 80% of the time, but also mis-classifies healthy patients 9.6% of the time.

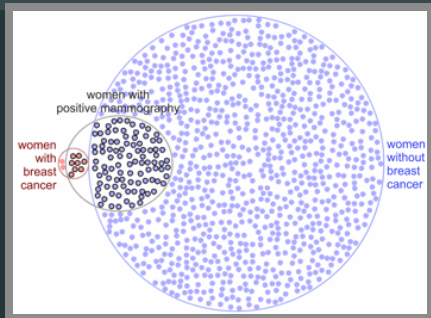
If a 40-year old woman has a positive mammogram test, what is the probability she has breast cancer?

Bayesian inference: example

Eddy's mammogram problem

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





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

$$\mathbb{E}[X] =$$

$$\mathbb{E}[g(X)] =$$

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

$$\mathbb{E}[X] =$$

$$\mathbb{E}[g(X)] =$$

variance and standard deviation

- **Definition:** $Var(X) =$ $\sigma(X) =$

- (More useful formula for computing variance)

$$Var(X) =$$

independence

what do we mean by “random variables X and Y are independent”?
(denoted as $X \perp\!\!\!\perp Y$; similarly, $X \not\perp\!\!\!\perp Y$ for ‘not independent’)

intuitive definition: knowing X gives no information about Y

formal definition:

- One measure of independence between rv is their covariance

$$\text{Cov}(X, Y) = \quad \quad \quad \text{(formal definition)}$$

$$= \quad \quad \quad \text{(for computing)}$$

independence and covariance

how are independence and covariance related?

- X and Y are independent, then they are **uncorrelated**
in notation: $X \perp\!\!\!\perp Y \Rightarrow \text{Cov}(X, Y) = 0$
- however, uncorrelated rvs can be dependent
in notation: $\text{Cov}(X, Y) = 0 \nRightarrow X \perp\!\!\!\perp Y$
- $\text{Cov}(X, Y) = 0 \Rightarrow X \perp\!\!\!\perp Y$ only for **multivariate Gaussian rv**
(this though is confusing; see [this Wikipedia article](#))

linearity of expectation

for any rvs X and Y , and any constants $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

note 1: **no assumptions!** (in particular, does not need independence)

linearity of expectation

for any rvs X and Y , and any constants $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

note 1: no assumptions! (in particular, does not need independence)

note 2: does not hold for variance in general

- for general X, Y

$$\text{Var}(aX + bY) =$$

- when X and Y are independent

$$\text{Var}(aX + bY) =$$

using linearity of expectation

the TAs get lazy and distribute graded assignments among n students uniformly at random. On average, how many students get their own hw?

using linearity of expectation

the TAs get lazy and distribute graded assignments among n students uniformly at random. On average, how many students get their own hw?

Let $X_i = \mathbb{1}_{[\text{student } i \text{ gets her hw}]}$ (indicator rv)

$N =$ number of students who get their own hw $= \sum_{i=1}^n X_i$

then we have:

$$\begin{aligned}\mathbb{E}[N] &= \mathbb{E}\left[\sum_{i=1}^n X_i\right] \\ &= \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \sum_{i=1}^n \mathbb{P}[X_i = 1] = \sum_{i=1}^n \frac{1}{n} = 1\end{aligned}$$

sums of independent rv

sums and averages of independent rv

- X_1, X_2, \dots are independent random variables that are uniformly distributed over the interval $[0, 1]$.
 - $\mathbb{E}[X_1] = 1/2$, $\text{Var}(X_1) = 1/12$.
 - The probability density function of X_1 looks like
-
-
-
-
-
-
-
-
-
-
- What about the pdf of $X_1 + X_2$ and $(X_1 + X_2)/2$?

sums and averages of independent rv

$$S_n = X_1 + \dots + X_n \qquad \bar{X}_n = \frac{1}{n} [X_1 + \dots + X_n]$$

- $\mathbb{E}[S_n] =$

$$\mathbb{E}[\bar{X}_n] =$$

$$\text{Var}(S_n) =$$

$$\text{Var}(\bar{X}_n) =$$

- (roughly) sum of n i.i.d. random variables is \sqrt{n} times as variable as any one of the random variables
- average of n i.i.d. random variables is $1/\sqrt{n}$ times as variable as any one of the random variables