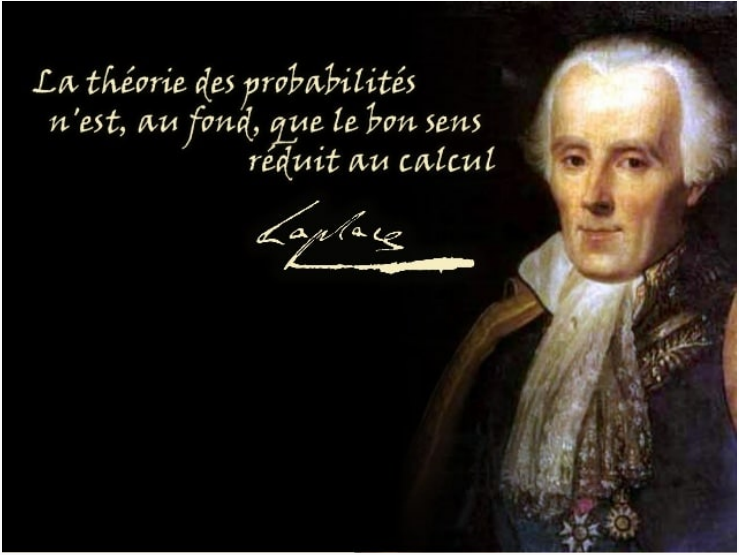
ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

unit 1: probability review

Sid Banerjee

School of ORIE, Cornell University



"probability theory is common sense reduced to calculation"

not quite...

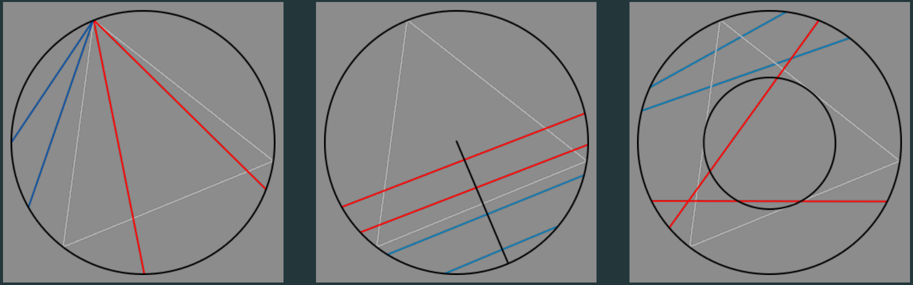
Bertrand's problem

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

not quite. . .

Bertrand's problem

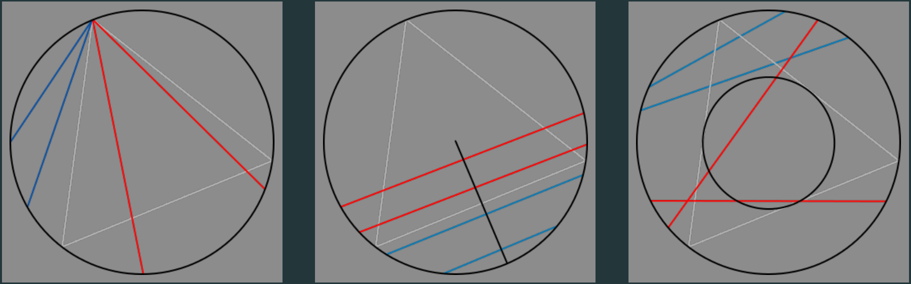
given equilateral triangle inscribed in a circle and a random chord C, what is the $\mathbb{P}$[length of C exceeds the side of the triangle]?



not quite...

Bertrand's problem

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



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 be comfortable with

- random variables (rv) + cumulative distribution fn (cdf)
- expectation and variance of random variables
- independence (and dependence – mutual exclusivity, 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 $\Omega \to \mathbb{R}$

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 (or so I sometimes feel...) 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

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 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?

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

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?

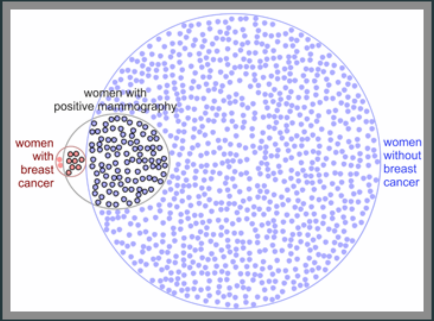
Bayesian inference: example

Eddy's mammogram problem

– $\mathbb{P}$[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?



credit: Micallef et al.

expectation 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

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

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

independence and covariance

how are independence and covariance related?

- X and Y are independent, then they are uncorrelated

  in notation: $X \perp\!\!\!\perp Y \quad \Rightarrow \quad Cov(X, Y) = 0$

- however, uncorrelated rvs can be dependent

  in notation: $Cov(X, Y) = 0 \quad \not\Rightarrow \quad X \perp\!\!\!\perp Y$

- $Cov(X, Y) = 0 \Rightarrow X \perp\!\!\!\perp Y$ 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

  $$Var(aX + bY) =$$

- when X and Y are independent

  $$Var(aX + bY) =$$

using linearity of expectation

the TAs get lazy and distribute graded assignments among n students u.a.r.

on average, how many students get their own hw?

using linearity of expectation

the TAs get lazy and distribute graded assignments among n students u.a.r.
on average, how many students get their own hw?

Let $X_i = \mathbb{1}_{[\text{student i gets her hw}]}$ (indicator rv)
$N =$ number of students who get their own hw $= \sum_{i=1}^{10} X_i$
then we have:

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

sums of independent rv

sums and averages of independent rv

- $X_1, X_2, \ldots$ are independent random variables that are uniformly distributed over the interval $[0, 1]$.
- $\mathbb{E}[X_1] = 1/2$, $Var(X_1) = 1/12$.
- 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

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

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

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

  $Var(S_n) =$

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

- (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