can (e.g.) hold β at 0.999 and γ at 0.994 and vary $\alpha = P(\theta = 1 \mid \mathcal{B})$ from (say) 0.005 to 0.02 (a factor of 2 lower & higher than the previous value of 0.01): 39

(β = 0.999, γ = 0.994)		
α	FPR	FNR
0.005	0.544	0.00000506
0.01	0.373	0.0000102
0.02	0.227	0.0000205

cutting the prevalence in half increases FPR by $\left| \frac{.544 - .373}{.373} \right| \cdot 100\% \doteq 46\%$; doubling the prevalence decreases FPR by $\left| \frac{.227 - .373}{.373} \right| \cdot 100\% \doteq 39\%$.

(15 Jun 19) (i.e., the false positive rate is quite sensitive to prevalence (prior information)).

Cutting α in half almost exactly cuts the FNR in half, and doubling α almost exactly doubles FNR, so the false negative rate is also quite sensitive to prevalence (although its value remains extremely low).

Probability Interlude

In CS 1 I used symbols such as $P(\theta = 1 \mid \mathcal{B})$ without defining $P(\cdot \mid \cdot)$, and I assumed some Axioms that become Theorems when deeper Axioms are adopted; let's dig deeper.

The first careful attempt to <u>define probability</u> occurred in <u>1654</u>, in an exchange of letters between Pascal & Fermat (P&F); they were motivated by <u>gambling questions</u> posed by their friend the <u>Chevalier de Méré</u>.

Example: which is a better bet: (At least 1 ace $\boxed{\cdot}$ in 4 rolls of a die) or (At least 1 double ace $\boxed{\cdot}\,\boxed{\cdot}$ in 24 rolls of a pair of dice)?

P&F began by assuming that what we would now call a <u>random experiment</u> can be characterized by its <u>elemental outcomes</u> (EOs): the set S (tacitly assumed finite) of all EOs is an approach to summarizing all the ways the experiment could come out in such a way that the EOs are <u>equipossible</u>, i.e., there is no reason that any EO should be favored over any other EO. <u>If</u> this can be done, then

$\boxed{\text{Def.}}$ $$P_{PF}(A) \stackrel{\Delta}{=} \frac{\#\text{ EOs favorable to } A}{\text{total } \#\text{ of EOs}};$$

(is defined to be)

here A is a set of EOs that is a subset of S.

Another name for P&F the assumed set up above is the Equally Likely model (ELM)

Toy example

population — sample

$$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow{\text{at random}} [I_1]$$

ELM? yes, by "at random"

so $P_{PF}(I_1 \text{ is odd}) = \frac{2}{3}$

In the problem formulation I introduced (42)

in CS1, $\mathcal{B} = \{ B_1 \}$, $B_1 = \left(\begin{array}{c} \text{(T/F proposition)} \\ I_1 \text{ chosen at random} \\ \text{from } \{1, 2, 4\} \end{array} \right)$,

and $P_{PF}(\underbrace{I_1 \text{ is odd}}_{\substack{\text{T/F proposition} \\ A}}) = P(A \mid \mathcal{B}) = \frac{2}{3}$.

Chevalier de Méré

$A_1 = ($at least 1 ace in 4 rolls of a die$)$

$P_{PF}(A_1)$ is undefined unless You assume ($\mathcal{B}$) that the die-rolling is fair: if so, all 6 possibilities

$\{1, 2, \dots, 6\}$ on a single roll are "equally probable" (P&F equipossible) and the outcomes of the 4 rolls are independent (information about the outcome of any single roll doesn't change the chances for any other roll).

Under this assumption the $6^4 = 1296$ possibilities $\{1111, 1112, \dots, 6666\}$ are equipossible (so the set of EOs is $S = \{1111, \dots, 6666\}$) 43

and careful counting reveals that 671 are favorable to A_1, so $P_{PF}(A_1) = \frac{671}{1296} \doteq 0.518$

My problem formulation plus AMS 131 ideas:

$P(A_1 \mid \mathcal{B})$, $\mathcal{B} = \{B_1, B_2\}$, $B_1 = \begin{pmatrix} \text{6 faces of} \\ \text{die equally} \\ \text{likely} \end{pmatrix}$,

← proposition; ← set of propositions

$B_2 = \begin{pmatrix} \text{4 rolls} \\ \text{independent} \end{pmatrix}$

$$P(A_1 \mid \mathcal{B}) = 1 - P\left(\begin{matrix} \text{not} \\ A_1 \end{matrix} \,\middle|\, \mathcal{B} \right)$$

$$= 1 - P\left(\begin{matrix} \text{0 aces} \\ \text{in 4 rolls} \end{matrix} \,\middle|\, \mathcal{B} \right) = 1 - P\left[\begin{pmatrix} \text{not} \\ \text{ace} \\ \text{on} \\ \text{1st} \\ \text{roll} \end{pmatrix} \underline{\text{and}} \dots \underline{\text{and}} \begin{pmatrix} \text{not} \\ \text{ace} \\ \text{on} \\ \text{4th} \\ \text{roll} \end{pmatrix} \,\middle|\, \mathcal{B} \right]$$

$$\overset{(B_2)}{=} 1 - \left\{ P\left[\begin{pmatrix} \text{not} \\ \text{ace} \\ \text{on} \\ \text{1st} \\ \text{roll} \end{pmatrix} \,\middle|\, \mathcal{B} \right] \cdot \dots \cdot P\left[\begin{pmatrix} \text{not} \\ \text{ace} \\ \text{on} \\ \text{4th} \\ \text{roll} \end{pmatrix} \,\middle|\, \mathcal{B} \right] \right\}$$

$$\overset{(B_1)}{=} 1 - \underbrace{\left(1 - \frac{1}{6}\right) \cdots \left(1 - \frac{1}{6}\right)}_{4} = 1 - \left(1 - \frac{1}{6}\right)^4 \doteq 0.518$$

$A_2 = \begin{pmatrix} \text{at least 1} \\ \text{double ace in} \\ \text{24 rolls of a pair of dice} \end{pmatrix}$

Under the fairness assumption about the dice-rolling, there are $36^{24} \doteq 2.25 \cdot 10^{37}$

EOs in $\mathcal{S} = \{ \underbrace{(1,1), \dots, (1,1)}_{\leftarrow 24 \rightarrow}; \dots; \underbrace{(6,6), \dots, (6,6)}_{\leftarrow 24 \rightarrow} \}$,

and "careful counting" reveals that $P_{PF}(A_2) =$

$$1-\left(1-\frac{1}{36}\right)^{24} \overset{\text{(maple)}}{=} \frac{11033126465283976852912127963392284191}{22452257707354557240087211123792674816} \text{ (!)}$$

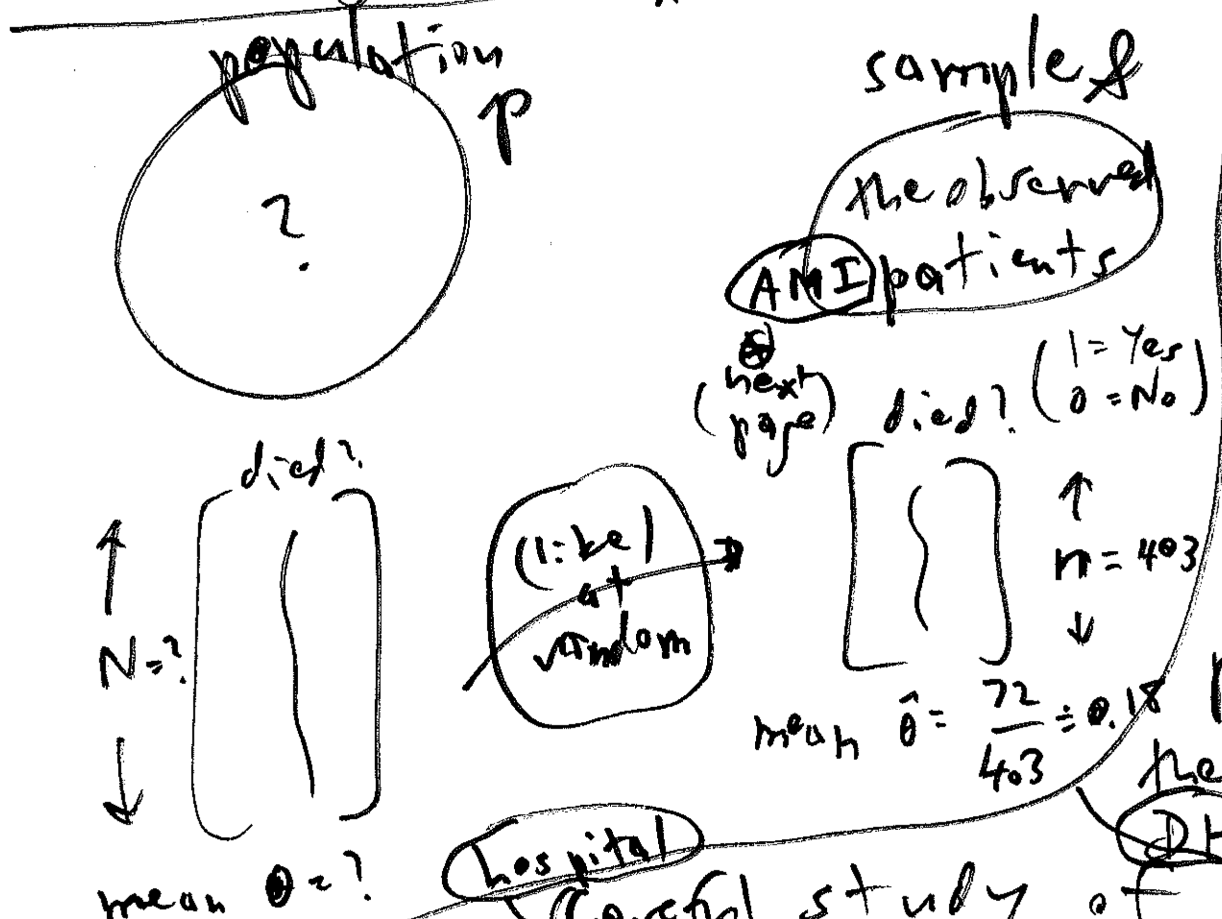
$\doteq 0.491$. So A_1 and A_2 are both roughly 50/50 propositions. Clearly the only way P&F could have computed $P_{PF}(A_2)$ was by working out the (AMS 131) rules for <u>and</u> and <u>not</u> (and certainly they must have understood <u>or</u> also). <u>Advantages</u> of the P&F approach: it's simple, when the ELM applies; ..., er, that's about it. <u>Disadvantages</u>:

① The P&F definition of probability is essentially circular: equipossible = equiprobable, and you can't use probability to define probability

---

② Lots of interesting problems involving uncertainty can't be forced into the "equipossible elemental outcomes" framework (e.g., what's the P&F probability that Shakespeare wrote everything we attribute to him?). ③ Doesn't immediately apply to situations in which the number of possible experimental outcomes is (countably or uncountably) infinite. P&F did not attempt to use probability to learn about the world from data; their use of probability was purely <u>deductive</u>; in contrast, our use of Bayes's Theorem in CS 1 was an example of <u>inductive</u> reasoning, which (when you're learning from data) is statistical <u>inference</u>

Basic diagram of (inferential) statistical learning:



Case study 2 | IQ: How good Is the quality of care for hospitalized heart attack patients at the Dominican Hospital (DH)? (Santa Cruz)

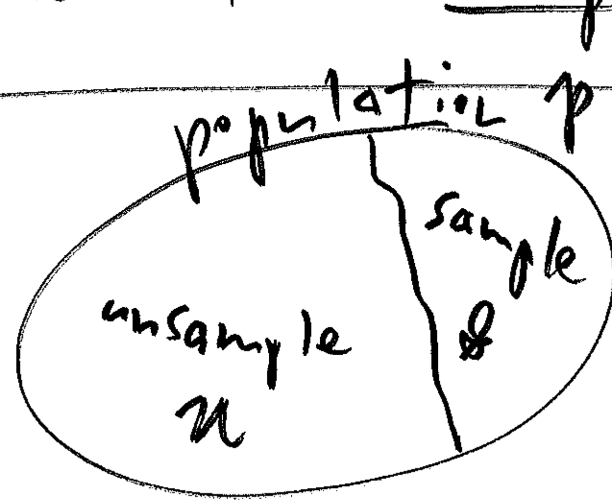
Careful study of quality of care involves gathering 4 kinds of data: <u>process</u> (what health care professionals do for patients); <u>outcome</u> (what happens as a result of process); <u>sickness at admission</u> (hospitals treating sicker patients are expected to have worse outcomes); and <u>structure</u> (does the DH have the latest scanning (imaging) technology?).

In CS2 we'll look at an extremely simplified version of this setup, based only on a single outcome (the "hardest endpoint" of all: 1 if patient died within 30 days of admission, 0 otherwise) & ignoring process, sickness & structure.

Fancy medical name for heart attack: acute myocardial infarction (AMI). (*)

One simple idea: open a time window, e.g., 1 Jan 2015 → 31 Dec 2018, look at records of all n AMI $\boxed{\text{Medicare}}$ (see p. 48) patients at DH in that time window (e.g., $n = 403$), count how many died (call this number $s = 72$ (say)), compute 30-day death rate $\hat{\theta} = \frac{s}{n} = \frac{72}{403} \doteq 0.18$, compare this with national 30-day death rate r_{AMI} for AMI Medicare patients in 2015-2018; if $\hat{\theta}$ is a lot bigger than r_{AMI}, this suggests a possible quality of care problem for treatment of AMI at DH ("prima facie" evidence). I focus here on <u>Medicare</u> patients (age 65+, or disabled, or both) because the link between process and outcome is stronger for them than for non-Medicare patients, making them better candidates for quality-of-care evaluation. For Medicare patients $r_{AMI} \doteq 0.132$ (Cedars Sinai hospital web site), a lot lower than $\hat{\theta} = 0.18$ (this is hypothetical, but if true ~~the~~ DH would have some explaining to do). I'm pretending here that I took a complete <u>census</u> of all Medicare AMI patients at DH in 2015-2018; $\hat{\theta} = \frac{s}{n} \doteq 0.18$ is a <u>description</u>

of their mortality experience. What if I wanted to *generalize outward* to other AMI patients from this data set? This involves imagining a larger world of patients: such a world is called a *population* $\mathcal{P}$. The $n = 403$ DH patients form a *subset* of $\mathcal{P}$; such a subset is called a *sample* $\mathcal{S}$ from $\mathcal{P}$. The presence of $\mathcal{S}$ *partitions* $\mathcal{P}$ into 2 non-overlapping & exhaustive subsets: the *sample* $\mathcal{S}$ and the *unsample* $\mathcal{N}$.



$\boxed{Q:}$ How should the sample be chosen so that it accurately reflects all of $\mathcal{P}$?

key idea | A: | Try to make the sample ~~S~~ and the unsample N as similar as possible in all relevant ways. | A sampling method that promotes this goal is said to be weak-sense representative of the population P.

Q: | How achieve this goal? | A: | The simplest approach is to choose the sampled patients at random from the population.

The idea dates back to biblical times, & Pierre-Simon (de) Laplace (1749-1827) estimated the population of France with random sampling in 1786; in the 20th century random sampling was forcefully advocated by Jerzy Neyman (1894-1981) in a paper in 1934 that laid the foundations for contemporary scientific sampling. (17 Jan 19)

Two simple ways to do random sampling: at random with replacement (independent identically distributed (IID) sampling) & at random without replacement (simple random sampling (SRS)).

Toy example

~~population~~ sample

$$N = 3 \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix} \xrightarrow{\text{at random}} \begin{array}{c} I_1 \\ I_2 \end{array} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad n = 2$$

sample size

population size

Capital

Let I_1, I_2 be random variables (AMS 131) recording the results (y_1, y_2) lower case of $n=2$ random draws from this toy population; $B_i = (B_i)$, $i=1,2$; $B_1 =$ (draws are IID); $B_2 =$ (draws are SRS)