Charles MARGOSSIAN

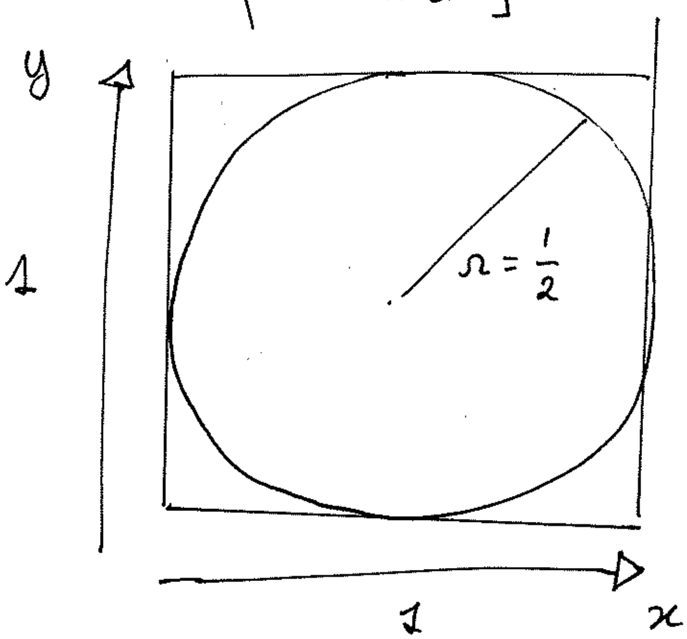
Biometry in Pharmaceutics
Spring 2022

Probability & Bayes I

1/ Buffon's Needle

In 1773, Buffon claims he can calculate π by dropping a needle.

[For this lecture, consider a simplified experiment.]



Setup: drop a needle "randomly" on a unit square and record the position where the needle falls.

Denote (x, y) the positions of the needle...

Question: what are the odds the needle falls in the circle?

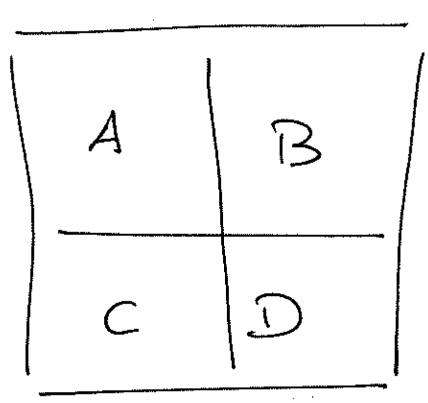
More

Questions: - Is this experiment practical?

- What conditions must, when dropping, the needle must be met?

→ Needle can hit any point with the same probability:

- $x \sim$ uniform $(0, 1)$
- $y \sim$ uniform $(0, 1)$



- Partition square in 4 equal squares
- What is the prob the needle falls in square A?

$$Pr(A) = \frac{1}{4}.$$

- If we partition the square into N equal squares, probability we fall into one particular square is

$$Pr(a) = \frac{1}{N}.$$

- Area of surface determines probability of falling in that surface.

Back to circle...

Area of the circle is

$$A_{\odot} = \pi r^2 = \pi \left(\frac{1}{2}\right)^2 = \frac{\pi}{4}.$$

Thus probability of falling in circle is $\frac{\pi}{4}$.

Idea: Estimate $Pr(\odot) = \frac{\pi}{4}$, and then get estimate for π.

Suppose we drop N needles.
Then an estimator for $Pr(\odot)$ is the fraction of needles which fall into the circles.

Question: how many needles do we need to drop?

Exercise: In R, simulate Buffon's experiment, and estimate $Pr(\odot)$.
Compare your estimate of $Pr(\odot)$ to the true value $\pi/4$ for varying values of N.
Does the accuracy of the estimator improve with N?

To study this question, consider the random variable

$$X = \begin{cases} 1, & \text{if needle falls in circle} \\ 0, & \text{otherwise.} \end{cases}$$

Probability mass function

We can associate a p.m.f with X:

$$Pr(X = 1) = p$$
$$Pr(X = 0) = 1 - p$$

More generally,

$$Pr(X = x) = p^x (1-p)^{(1-x)}$$

(Bernoulli dist.)

Exercise: Show that the above ~~examp~~ formulation is consistent with the case where $X = 1$ and $X = 0$.

Other example: Poisson distribution.
X can be any integer, $0, 1, 2, \ldots, +\infty$, and $P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

Property: summing the p.m.f over all possible outcomes returns 1.

Example: (Bernoulli)

$$\Pr(X = 1) + \Pr(X = 0) = p + 1 - p = 1.$$

Example: (Poisson)

$$\sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1,$$

where we use a Taylor expansion of the exponential e^{λ}.

- Distributions also admit an expectation value, also termed the mean.
  - ↳ this is the sum of all possible outcomes weighted by their probability mass.

Example (Bernoulli)

$$\begin{aligned} \mathbb{E}X &= (1)\Pr(X=1) + (0)\Pr(X=0) \\ &= (1) \cdot p + (0) \cdot (1-p) = p \end{aligned}$$

Can interpret this through the lens of a population mean.

Suppose $p = 0.7$.

→ corresponds to a hypothetical population where 70 individuals have $X = 1$ dollars, and 30 individuals have $X = 0$ dollars.

Then, <u>on average</u>, each individual has 0.7 dollars.

<u>Example</u> (Poisson)

$$\mathbb{E} X = \lambda \qquad \text{(proof omitted)}$$

Suppose, in Bernoulli example, we do not survey all 100 people, but only 10 people. How can we estimate the population mean?

→ Sample mean: $\hat{p} = \frac{1}{10} \sum_{i=1}^{10} x_i$

- Is this estimator unbiased?

  ↳ assume $X_1, X_2, \ldots, X_{10}$ are independent and identically distributed (i.i.d).

<u>Property</u>: If $X_1, X_2, \dots, X_n$ are i.i.d, with expectation value μ, then

$$\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mathbb{E} X_i = \frac{1}{N} N\mu = \mu .$$

$\Rightarrow$ Sample mean is <u>unbiased</u>.

<u>Variance</u>:

The mean tells us what happens on <u>average</u>. The <u>variance</u> measures how much we might expect variables to deviate from the mean, i.e. their average behavior.

<u>Example</u>:

- If the temperature is 30°F every day, the mean temperature is 30°F and the variance is 0.
- If the temperature is 0°F one day and 60°F the next day, the mean temperature is 30°F but the variance is $(30°F)^2$.

Definition :

$$\mathrm{Var}\, X = \mathbb{E}\left((X - \mathbb{E}X)^2\right).$$

The standard deviation is $SD = \sqrt{\mathrm{Var}\, X}$.

Property : $\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2$

Exercise : In the Bernoulli case, show that

$$\mathrm{Var}\, X = p \cdot (1 - p).$$

Remark : when $p = 0$ or $p = 1$, the variance is 0. Does this make sense?

Exercise : Show that the variance is maximized for $p = \frac{1}{2}$.

Property : Let $a \in \mathbb{R}$ (i.e. a is constant). Then

$$\mathrm{Var}(aX) = a^2\, \mathrm{Var}(X).$$

Exercise : Prove the above equation.

A consequence of the above result is that

$$\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)$$

$$= \frac{1}{N^2}\,\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right)$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}\,X_i \qquad \text{(by independence of the } X_i\text{'s)}$$

$$= \frac{1}{N^2}\,N\,\mathrm{Var}\,X$$

$$= \frac{1}{N}\,\mathrm{Var}\,X.$$

<u>Question</u>: what happens as N grows?

In the limit,

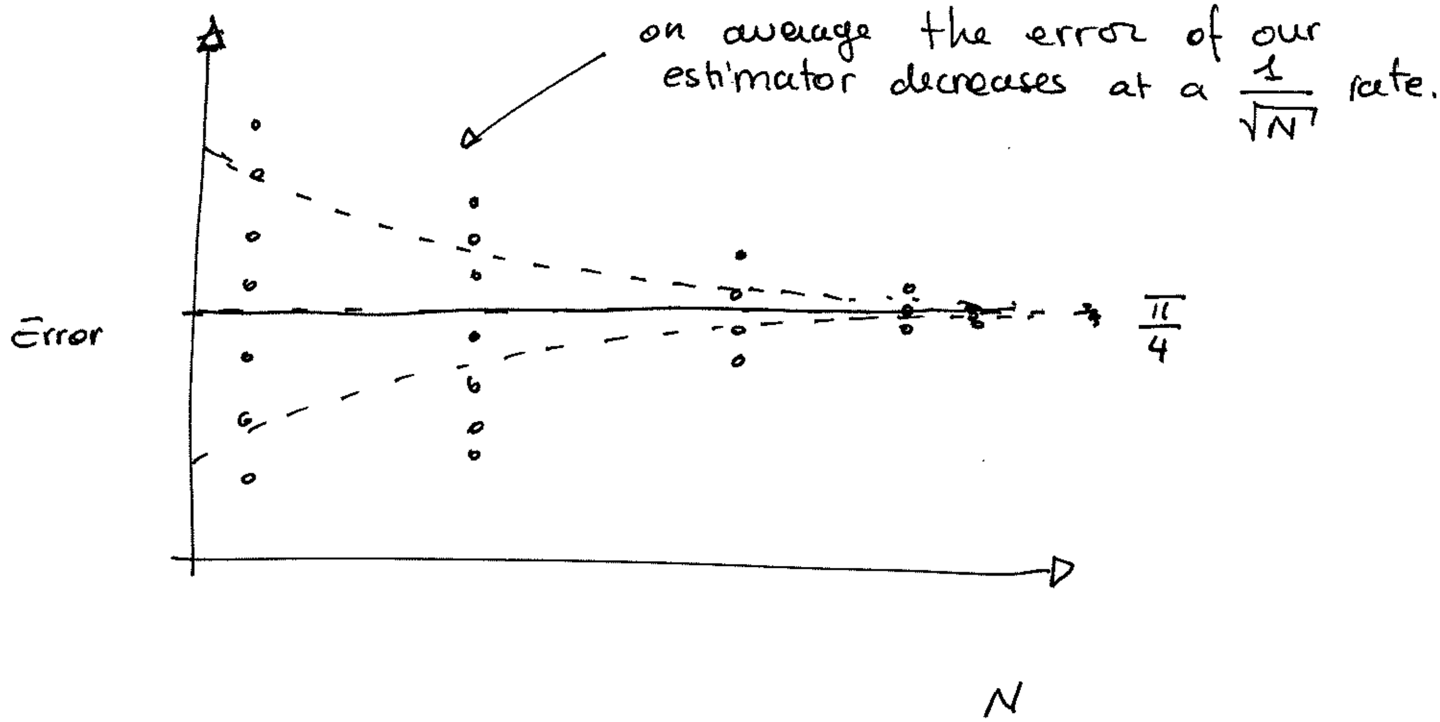
$$\lim_{N\to+\infty}\mathrm{Var}(\bar{X}) = 0,$$

where $\bar{X} = \frac{1}{N}\sum X_i$.

Combined with the unbiasness property, we have that the sample mean, $\bar{X}$, is ("with high probability") equal to $\mathbb{E}X$, when $N \to +\infty$.

This the <u>Weak Law of Large Numbers</u>.

Suppose we repeat Buffon's experiment many times, for varying number of trials, N, and compute the error of our estimate.



<u>Exercise</u>: In Buffon's experiment, show that for

$$X = \begin{cases} 1 & \text{if Needle falls in circle} \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{Var } X = \frac{3}{16}\pi^2.$$

Therefore $\text{Var } \bar{X} = \frac{1}{N}\frac{3}{16}\pi^2$,

and $SD(X) = \frac{1}{\sqrt{N}}\frac{\sqrt{3}}{4}\pi$.

How does this compare to your simulations in R?

Conditional Probability

- Consider two random variables X and Y. Does knowing the value X teach me something about the potential value of Y? If so, the variables are dependent.

Example:

X : the weather in the morning in Bufallo

Y : the weather in the afternoon in Bufallo.

The two events are dependent.

Example 2:

X : the weather in the morning in Bufallo.

Y : the weather in the afternoon on Mars.

The two events are independent?
(Depends on your model of weather)

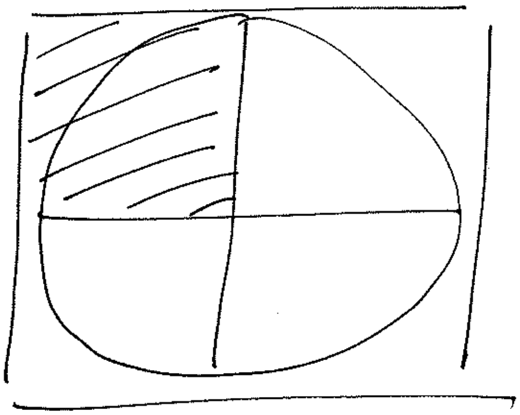
Example 3:

X : the position of the first needle.

Y : the position of the second needle.

→ Independant.

Do example 5 first during lecture ...



Example 4*:

X : the needle falls in the circle.

Y : the needle falls in the shaded square.

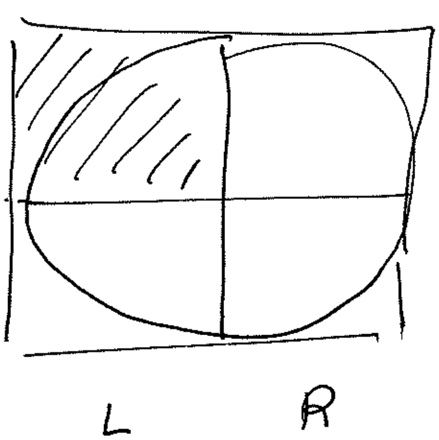
$$P(X) = \frac{\pi}{4} \quad \text{and} \quad P(X \mid Y) = \frac{\pi/4\,/4}{1/4} = \frac{\pi}{4} = P(X)$$

this is the prob of X given Y.

$$P(Y) = \frac{1}{4} \quad \text{and} \quad P(Y \mid X) = \frac{1}{4} = P(Y)$$

$\Rightarrow$ The two events are <u>independent</u>!

Example 5*:



X : the needle falls in the left side.

Y : the needle falls in the shaded square.

$$P(Y) = \frac{1}{4}, \quad P(Y \mid X) = \frac{1}{2}$$

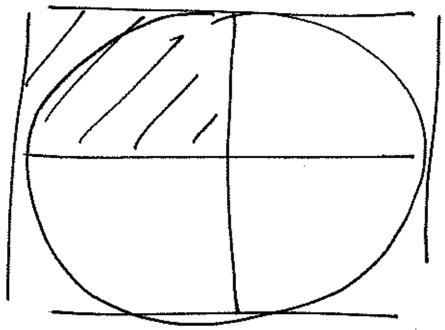
$$P(X) = \frac{1}{2}, \quad P(X \mid Y) = 1.$$

$\Rightarrow$ Events are <u>dependent</u>.

Joint distribution / Probability

We can ask what is the probability that two events happen conjointly.

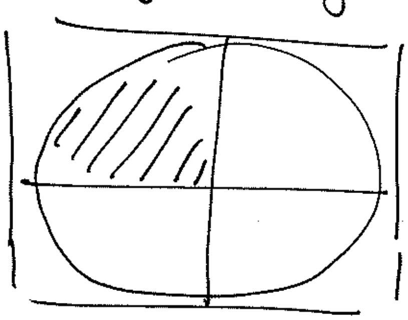
Example :



X : the needle falls in the circle.

Y : needle falls in shaded square.

The probability that X and Y happen simultaneously is given by the shaded area below.



$$P(X, Y) = \frac{1}{4}\frac{\pi}{4} = \frac{\pi}{16}.$$

Property : $P(X,Y) = P(X)\,P(Y|X)$.

In example 4 (above),

$$\boxed{P(X,Y) = P(X)\,P(Y)} = \frac{1}{4}\frac{\pi}{4} = \frac{\pi}{16}.$$

this is the case when X and Y are independent.

In example 5, $P(X,Y) = P(X)\,P(Y|X)$

$$= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} \neq P(X).P(Y).$$

Bayes' rule:

$$P(X \mid Y) = \frac{P(Y \mid X)\, P(X)}{P(Y)}$$

Proof:

$$P(Y \mid X)\, P(X) = P(Y, X).$$
$$= P(Y)\, P(X \mid Y)$$

Then

$$\frac{P(Y \mid X)\, P(X)}{P(Y)} = \frac{P(Y)\, P(X \mid Y)}{P(Y)} = P(X \mid Y) \quad \blacksquare$$

Example (Mammogram)

Interested in test with a certain error rate.

Test \ Truth	B	$\bar{B}$
A	0.9	0.1
$\bar{A}$	0.2	0.8

$A = \{\text{Test is positive}\}$

$B = \{\text{Patient has cancer}\}$

Denote $\bar{A}$ the complementary event, i.e. patient has negative test. Same with $\bar{B}$.

False positive rate: $P(A \mid \bar{B}) = 0.1$

False negative rate: $P(\bar{A} \mid B) = 0.2$

Question: Given that we have a positive test, what is the probability of having cancer? i.e.

$$P(B \mid A) \; ?$$

Let's use Bayes' rule:

$$P(B \mid A) = \frac{P(A \mid B)\, P(B)}{P(A)}.$$

The true positive rate, $P(A \mid B)$, is

$$P(A \mid B) = 1 - P(\bar{A} \mid B) = 1 - 0.2 = 0.8.$$

Next the positive rate is

$$P(A) = P(A \mid B)\, P(B) + P(A \mid \bar{B})\, P(\bar{B}).$$

We are missing one piece of information to answer the question: the rate of individuals with cancer.

Let's look it up ... $P(B) = 0.04$.

Then $P(\bar{B}) = 0.96$.

Thus

$$P(A) = P(A|B)\,P(B) + P(A|\bar{B})\,P(\bar{B})$$

$$= (0.8)(0.04) + (0.1)(0.96)$$

$$= 0.128.$$

Thus

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

$$= \frac{(0.8)(0.04)}{0.128} = 0.25.$$

Lower than expected!!
How can we make sense of this?

Suppose there 1,000 people:

- 40 have cancer, 960 don't. All get tested.
- Among the people with cancer,
  8 test negative, 32 test positive.
- Among those without cancer,
  96 test positive, 864 test negative.

$\Rightarrow$ The population of people who test positive is dominated by patients without cancer.

||