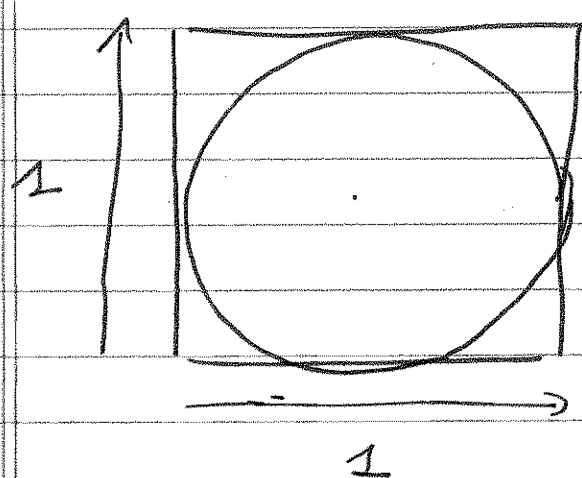
Biometry in Pharmaceutics
Probability & Bayes

U. Buffalo, March 6th

I/ Buffon's needle

For this lecture, we'll consider a simplified needle experiment.
The goal is to measure π.
Originally proposed in 1733.

Simplified Experiment



What are the odds the needle falls in the circle?

$p = \pi r^2 / A^2 = \pi / 4$

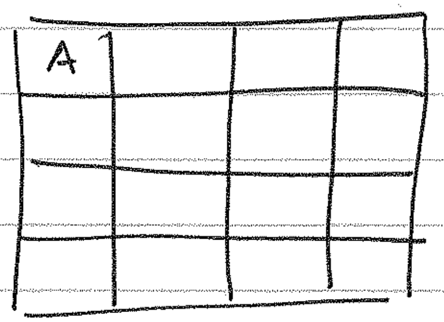
- Is this experiment practical?
- What conditions, when dropping the needle, must be met?
  - → Needle can hit any point with the same probability.

$x \sim U(0, 1)$

$y \sim U(0, 1)$

What are the odds the needle hits a specific point? Say (x_0, y_0)?

$\rightarrow$ simpler problem:



What are the odds the needle falls in box A?

$$p(A) = 1/16 = 1/N$$

where N is the total number of outcomes.
But a point has dimension 0.
Thus

$$P_r(x = x_0, y = y_0) = 0$$

This does <u>not</u> mean the event is impossible!
Hints at some counter-intuitive results when we deal with continuous spaces.

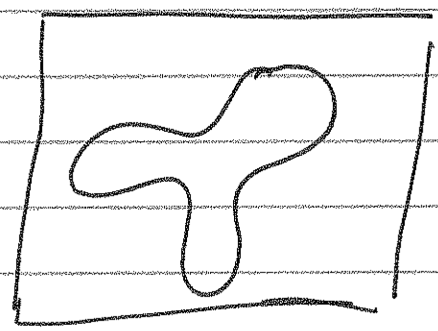
<u>Probability measure</u>: maps a set to $[0;1]$.

Example of sets/events

	<u>measure</u>
◦ a point (x_0, y_0)	0
◦ the box A	$1/16$
◦ the circle	$\pi/4$
◦ the whole space	1

Our prob. measure is the <u>area</u> of the set.

Suppose π is unknown (which it is!)
Or suppose we have a complicated shape:



How can we measure the area? Using probability.

define random variable

Let $A = \{$needle falls in circle$\}$
A^c the complement.
Let $p = Pr(A)$.
Then $X = 1$ or 0 (success / failure)
Suppose we drop n needles.
We say

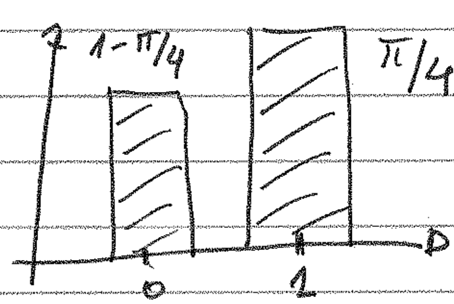
$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$$

Talk about density + prob. mass

$$\beta \propto p^{x}(1-p)^{1-x}$$

Talk about independence

Assigns probability mass to $x=1$ and $x=0$



If we drop a lot of needles, the average of X converges to p!

STRONG LAW OF LARGE #!

$\bar{X} \longrightarrow E\{X\}$ | roughly speaking

Expectation value: | $\lim_{n \to \infty} \frac{1}{n} \sum x_j = \mu$ a.s.

$\mu = E\{X\} = 1\, P(X=1) + 0 \cdot P(X=0) = p$ w/ probability 1.

~~WEAK LAW OF LARGE NUMBERS~~

$\cancel{\lim_{n \to +\infty} \Pr(|\bar{X}_n - \mu| > \epsilon) = 0}$, ~~for an arbitrarily small~~ $\cancel{\epsilon}$.

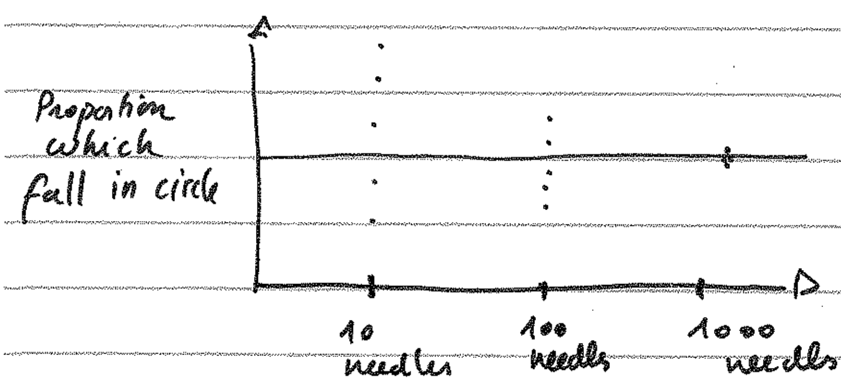
Central Limit Theorem

Tells us something about the rate of convergence.

How much variance will I have?

$$(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2/n)$$

$$\Leftrightarrow \quad \bar{X}_n \xrightarrow{d} N(\mu, \sigma^2/n)$$



- Need to define (a) variance

(b) Normal density

<u>Variance of Bernoulli</u>

$$\mathrm{Var}[X] = E[X^2] - E[X]^2$$
$$= \Pr(X=1)\, 1^2 = p$$
$$\Rightarrow v = p - p^2 = p(1-p). \ \square$$

$$v = p(1-p)$$

$\rightarrow$ compare $p = 1/2$

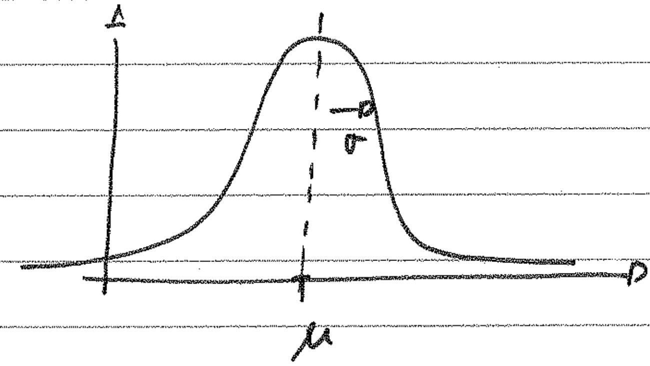
$p = 3/4$

$p = 1$

$p = 0$

<u>Normal density</u>

$$N(\theta, \sigma)$$



Introduce CDF

$$p_{\theta,\sigma} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \text{ for } X \sim N(\theta, \sigma^2)$$

What if we generate N iid X ?

$$X_1, \dots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$$

$$p_{\theta,\sigma}(X_1, \dots, X_n) = \prod_{i=1}^{n} p(X_i) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}\sum (x_i - \mu)^2}$$

Independent probabilities multiply! Use Bernoulli example + construct Binomial

What about expectations?

For discrete distribution $\mu = \sum_i x_i \Pr(X = x_i)$

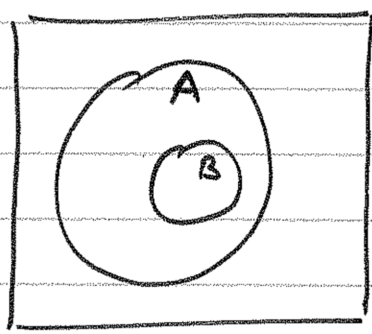
For continuous distributions:

$$\mu = \int_{\mathbb{R}} x \, p_{0,\sigma} \, dx$$

More generally, for any random function $f(X)$:

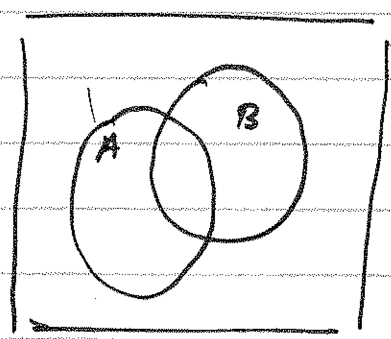
$$E\{f(X)\} = \int_{\mathbb{R}} f(x) \, p_{0,\sigma}(x) \, dx$$

Conditional Probability



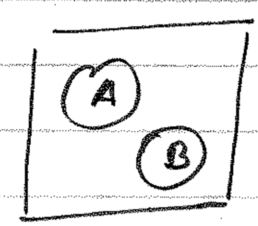
$$P(B|A) = \frac{P(B)}{P(A)}$$

What about $P(A|B)$?



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

What about



?

Baye's Theorem

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

Proof:

$$P(A \cap B) = P(A)\,P(B|A)$$
$$= P(B)\,P(A|B)$$

Thus

$$P(A)P(B|A) = P(B)P(A|B)$$
$$\Leftrightarrow \quad P(B|A) = \frac{P(B)\,P(A|B)}{P(A)} \qquad \blacksquare$$

Example: Mamograms

Let $B = \{$ patient has breast cancer $\}$
$A = \{$ Test is positive $\}$

$\Pr(A|B) = 0.8$ and $\Pr(A^c|B) = 0.2$
$\Pr(A|B^c) = 0.1$ and $\Pr(A^c|B^c) = 0.9$

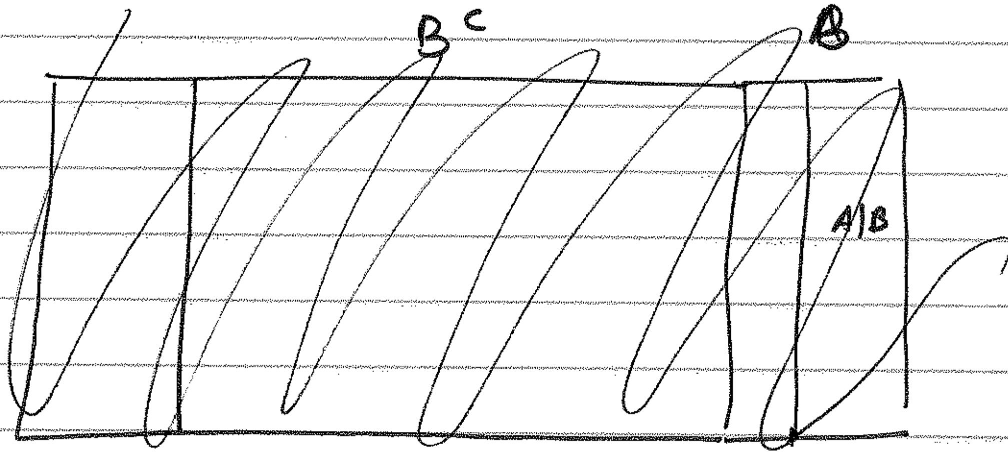
	+	−
cancer	0.8	0.2
no-cancer	0.1	0.9

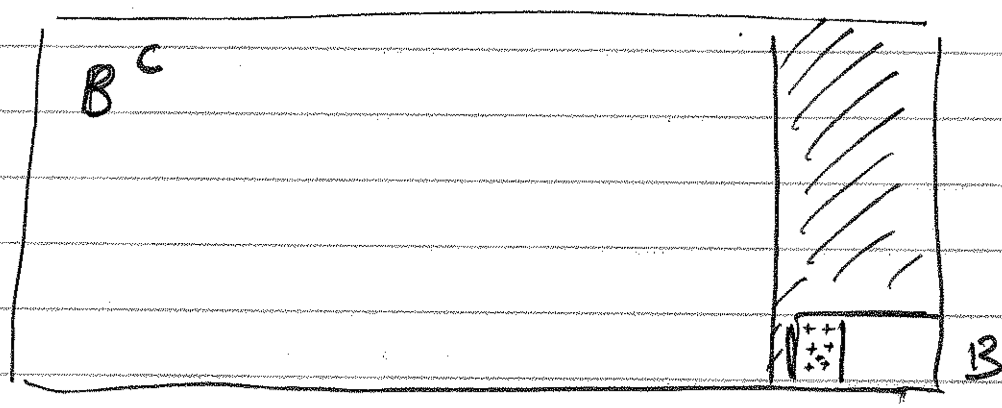
$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

$P(A) = P(A|B) + P(A|B^c)$
and $P(B) = 0.04\,\%$

Thus $P(B|A) = 0.3\,\%$ approx.

__Geometric Intepration__





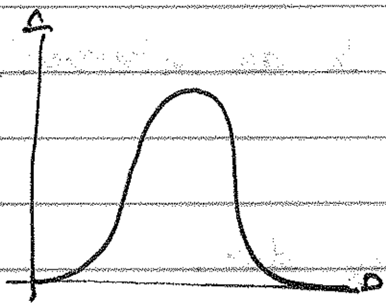
A patient is much more likely to land in the area, than in the . This is what Baye's rule tells us.

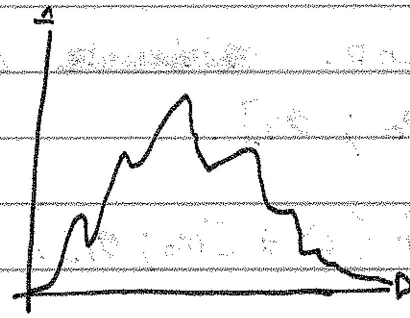
II/ Bayesian Inference

A goal in statistical inference is to characterize a DATA GENERATING PROCESS.

Note: some of statistics focuses on emulating the DGP to make prediction.
Tension between explain and predict.

Let $f(\cdot)$ be the density which generates data.
Can parametrize the density.





GAUSSIAN: μ, σ control this distribution.

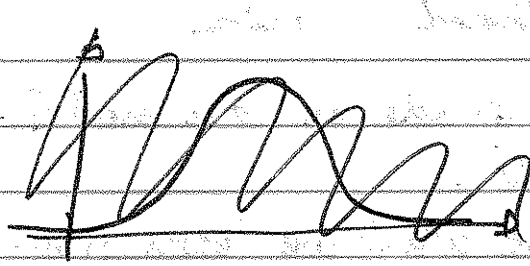
NON-PARAMETRIC
→ Infinite number of parameters.

In practice θ is unknown.
Can use $X_1, \ldots, X_n$ to estimate θ.
But how good is our estimate?

Explore a parameter space!!

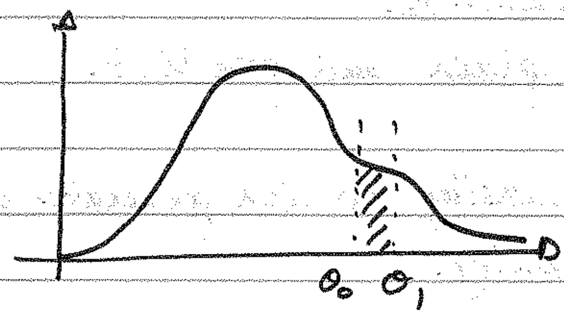
Random process ⇒ estimate is <u>not</u> perfect.

<u>Bayesian Idea</u>: use the posterior distribution of $\theta | X$ to describe our estimate $\hat{\theta}$.



$p(\theta|X)$

What can we learn from the posterior?



1/ Prob. ~~estimate~~ θ falls in an interval $[\theta_0 ; \theta_1]$.

$$Pr(\theta \in [\theta_0; \theta_1]) = \int_{\theta_0}^{\theta_1} f_{\theta|x} \, d\theta .$$

2/ Expected value of θ / mean value

$$E\{\theta\} = \int_{\Theta} \theta \cdot f_{\theta|x}(\theta) \, d\theta$$ ← not the same as the posterior mode!!

and similarly variance quantiles.

Computing the posterior

$$p(\theta|x) \propto \underbrace{p(x|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{Prior}}$$

- the likelihood depends on our model.
- so does our prior.

→ Interpretation: combine what we learn from the data with prior knowledge about θ.

The prior : - Prior knowledge :

- π is around 3.14
- A parameter has a mechanistic interpretation & can use expert knowldge.

E.g

V of gut. $\rightarrow$ w/ some variation.

$$V \sim N(V_0, \sigma_0^2)$$

- A regularization device
  - $\rightarrow$ can help model fitting process
  - $\rightarrow$ makes our parameter search more targeted exploration of target space.

E.g Immitation game / Code Breaking.

Example:

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2), \quad \sigma \text{ is known.}$$
$$\theta \sim N(\mu, \tau^2)$$

This prior and this likelihood play well together.

$$p_{X|\theta} = \frac{1}{\sqrt{2\pi}\,\sigma^{n}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

$$p_\theta = \frac{1}{\sqrt{2\pi}\,\tau} e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}$$

Produces a posterior $N\left(\dfrac{\mu/\tau + \sum x_i/\sigma^2}{1/\tau^2 + n/\sigma^2}, \dfrac{1}{1/\tau^2 + n/\sigma^2}\right)$

$$\Leftrightarrow \quad N\left(\frac{\mu/\tau^2 + \bar{X}/\sigma^2/n}{1/\tau^2 + n/\sigma^2}, \ldots\right)$$

The mean is a compromise or a weighted average between μ and $\bar{X}$.
The weighting is $1/VAR$

$\Rightarrow$ More uncertainty is penalized.

As $n \to +\infty$, the data/likelihood dominates the prior.

$$Var = \frac{1}{1/\tau^2 + n/\sigma^2}$$ goes to 0, if we have strong prior (low variance) or a lot of data.

OPEN UP FOR Questions, time allowing.

Further topics

- In practice need to estimate posterior distribution.
  - $\to$ MCMC etc.