The Last Lecture

Bayesian Statistics

→ In general, the analysis of almost all experiments that have ever been done has used the frequentist interpretation of statistical probability.

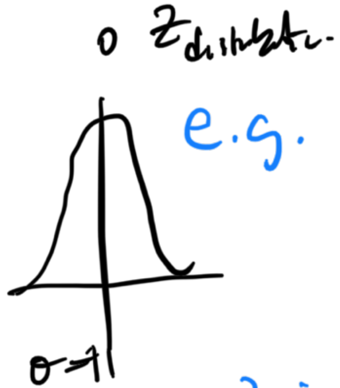
$$P_{\text{event}} = \text{limit of the frequency of an event after many trials } (n \to \infty)$$

→ In Bayesian statistics,

→ In Bayesian

P_{event} = degree of belief in the event happening

→ In Bayesian statistics, we compute and update probabilities after obtaining new data!



e.g.

$$\boxed{Z_{sample} = \frac{\bar{x} - \mu}{\sigma}}$$

→ in frequentist statistics, μ and σ are constants. In Bayesian statistics, they are not.

Mathematically and computationally

intensive !!

→ In 2020, this is not such a problem "☺"

→ See, for example, the Markov Chain Monte Carlo Algorithm.

Bayes' Theorem.

$$\underbrace{P(A|B)}_{\text{conditional probability}}^{\boxed{\text{analysis}}} = \frac{P(B|A)\,\underline{P(A)}}{\underline{P(B)}}$$

theory

$A \rightarrow$ proposition (e.g. a coin lands on heads 50% of the time)

"theory"
a priori

$B \rightarrow$ evidence

"data"

a posteriori

(e.g. the result of a series of coin flips)

$P(A) \rightarrow$ prior probability

→ expresses one's $\boxed{\text{beliefs}}$ about A before data is taken (prior knowledge)

P(B given A)

$P(B|A) \rightarrow$ $\boxed{\text{likelihood function}}$

→ the probability of obtaining the evidence, B, given that A is true.

→ Quantifies the extent to which the evidence, B, supports the proposition, A.

$P(A|B) \to$ posterior probability

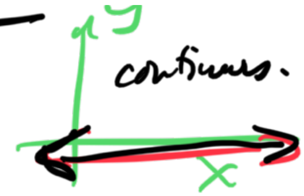
data analysis. $\to$ The probability that A is true, after taking the evidence into account

Bayes' theorem $\to$ we should update our prior beliefs after taking the evidence into account.

Why is it hard to calculate?

$$P(B) = \frac{P(B|A_1)\,P(A_1)}{+ P(B|A_2)\,P(A_2)}$$

$+$



$$+ \underline{P(B|A_n)}\ \underline{P(A_n)}$$

If there are an infinite number of possible outcomes, A_n, this involves integrals which are hard to calculate.

In practice, we often just say

$$\boxed{P(A|B) \propto \boxed{P(B|A)}\ P(A)},$$

and then use MCMC methods to estimate the constant of proportionality that optimizes the final result.

Practical Examples.

Vaccine Effectiveness

Prior Knowledge →

① vaccine is said to be "90% effective"

② 95% of the population gets vaccinated.

<u>Question</u> : What is the probability that a person who gets the disease was vaccinated?

$\boxed{10000}$ people.

× $P(B)$ $P(\bar{B})$

fewer | more

9500 vaccinated | 500 not vaccinated.

$P(A|B)$ ← "90% effective"

$P(A|\bar{B})$ ←

$9500(0.01) = \boxed{950}$ get disease. ↑ decrease

$\boxed{500}$ get disease. ↑ increase.

$$P(\text{vaccinated} \,/\, \text{get disease})$$

$$= \frac{950}{950 + 500}$$

$$= \underline{0.655} \quad !!$$

$(\underline{65.5\%})$

"90% effective"

→ So, about $\underline{2/3}$ of the people that get the disease were actually vaccinated!

Bayes' theorem

Problem

Psychological

A → get disease ✓

B → vaccinated ✓

$P(B) = 0.95$ ✓

$\boxed{P(B|A)}$ → probability that you were vaccinated, given that you got the disease.

Find.

$P(A|B)$ → probability that you got the disease, given that you were vaccinated.

$= 0.10$

$P(A)$ → probability that you get the disease (independent of whether you were vaccinated or not).

thing we want to find!

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

✓ reason.

$$\boxed{P(B|A)} = \frac{\overset{0.10}{\boxed{P(A|B)}} \cdot \overset{0.95}{\boxed{P(B)}}}{P(A)}$$

$$\boxed{P(A)} = \underbrace{P(A|B)}_{0.10}\,\underbrace{P(B)}_{0.95} + \underbrace{P(A|\bar{B})}_{1}\,\underbrace{P(\bar{B})}_{0.05}$$

$$P(B|A) = \frac{(0.10)(0.95)}{(0.10)(0.95) + (1)(.05)}$$

stupid people. → $= 0.655$ ☺

What about a $\boxed{99\%}$ effective vaccine?

$$P(B|A) = \frac{(0.01)(.95)}{(.01)(.95) + (1)(.05)}$$

$$= \boxed{0.1597}$$

(Still 16%!
1 in 6

What about if only $\boxed{90\%}$ are vaccinated.

$$P(B|A) = \frac{(0.01)(.90)}{(0.01)(.90) + 1(.10)}$$

!! → $= \boxed{0.083}$ ←

99.6% EF

95% V

Does that result surprise you?

Log Likelihood

Imagine an experiment with 4 possible

out comes. $(1, 2, 3, 4)$. You do the experiment a bunch of times, and get the following results:

$p \to$ heads.

$n_1 = 15$ ✓

$n_2 = 26$ ✓

$n_3 = 24$ ✓

$n_4 = 18$ ✓

$(HH) \to p^2$

$(HT) \to pq = p(1-$

$(TH) \to qp = p(1-p$

$(TT) \to q^2 = (1-p)^2$

83

Basic Probability of event $\to$ $\boxed{p \quad (=\tfrac{1}{2})}$

$$\boxed{\mathbb{P}_{\text{experiment}} = p_1^{n_1} \cdot p_2^{n_2} \cdot p_3^{n_3} \cdot p_4^{n_4}}$$

$$= (p^2)^{n_1} \cdot (pq)^{n_2} \cdot (qp)^{n_3} \cdot (q^2)^{n}$$

$$\mathbb{P} = p^{2n_1} \cdot p^{n_2}(1-p)^{n_2} \cdot (p^{n_3})(1-p)$$

$$+ (1-p)^{2n_4}$$

$$\boxed{P = \left[p^{\underline{2n_1 + n_2 + n_3}}\right]\left[(1-p)^{\underline{n_2 + n_3 + \ldots}}\right]}$$

maximum likelihood

$$\log(P) = (2n_1 + n_2 + n_3)\log(p)$$
$$+ (n_2 + n_3 + 2n_4)\log(1-p$$

$$\frac{d \log P_{exp}}{dp} = (2n_1 + n_2 + n_3)\left(\frac{1}{p}\right)$$
$$+ (n_2 + n_3 + 2n_4)\left(\frac{-1}{1-p}\right)$$

$$= 0$$

$$\frac{2n_1 + n_2 + n_3}{p} = \frac{n_2 + n_3 + 2n_4}{1-p}$$

$$2n_1 + n_2 + n_3 = p(n_2 + n_3 + 2n_4$$
$$+ 2n_1 + n_2 + n$$

$$= p(2n_1 + 2n_2 + 2n_3$$

$\boxed{\text{Log Likelihood}}$

$+2n_4)$

χ^2-mom.

$$p = \frac{2n_1 + n_2 + n_3}{2n_1 + 2n_2 + 2n_3 + 2n_4}$$

$$\boxed{p = \frac{2n_1 + n_2 + n_3}{2N}}$$

$$= \frac{2(15) + 26 + 24}{2(83)}$$

$$\boxed{p = 0.4819}$$

Update our belief

(i.e. NOT 50%)

Is this the correct model?

Calculate χ^2!

$$\chi^2 = \frac{(n_1 - n_1^{theory})^2}{n_1^{theory}} + \frac{(n_2 - n_2^{theory})}{n_2^{theory}}$$

$$+ \frac{(n_3 - n_3^{theo})^2}{n_3^{theo}}$$

$$+ \frac{(n_4 - n_4^{theo})^2}{n_4^{theo}}$$

$$n_1^{theo} = N \cdot p^2 = 19$$

$$n_2^{theo} = N \cdot pq = 21$$

$$n_3^{theo} = N \cdot qp = 21$$

$$n_4^{theo} = N \cdot q^2 = 22$$

$$\chi^2 = \frac{(15-19)^2}{19} + \frac{(26-21)^2}{21}$$

$$+ \frac{(24-21)^2}{21} + \frac{(18-22)^2}{22}$$

← not great!!

$$= \boxed{3.65}$$

Is there a better model?

The data is actually from the results of the NHL Stanley Cup final series over the last 83 years! ☺ It represents the number of games played before a winner was declared (4, 5, 6, or 7

Theory:

let p = probability that the "better" team wins a game.

q = probability that the "poor

team wins a game.

$\binom{5}{3} = \frac{5 \cdot 4}{2} = 1$

$\binom{6}{3} = \frac{{}^{2}\cancel{6} \cdot 5 \cdot \cancel{4}}{\cancel{3} \cdot \cancel{2}} = 20$

$$P_4 = p^4 + q^4$$

$$P_5 = \binom{4}{3}\left(p^3 q p + q^3 p q\right)$$

$$= 4\left(p^4 q + q^4 p\right)$$

$$P_6 = \binom{5}{3}\left(p^3 q^2 p + q^3 p^2 q\right)$$

$$= 10\left(p^4 q + q^4 p\right)$$

$$P_7 = \binom{6}{3}\left(p^3 q^3 p + p^3 q^3 q\right)$$

$$= 20 p^3 q^3 (\cancel{p + q}) = 20 p^3 q^3$$

$$P = \left(p^4 + q^4\right)^{n_1} \cdot \left(4\left(p^4 q + q^4 p\right)\right)^{n_2}$$

$$\cdot \left(10\left(p^4 q + q^4 p\right)\right)^{n_3} \cdot$$

$$\left(20 p^3 q^3\right)^{n_4}$$

Hard to solve, even with log likelihood!

So, solve it numerically!

Result: $p = 0.6541$

$\chi^2 = 0.363$

n_1	n_2	n_3	n_4	
15	26	24	18	data
[illegible]	24.1	23.3	19.2	theory