08/05/19

AMS 131

Lecture 4

$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ bit (John Tukey) 16 32,64 $1 \longleftrightarrow 1$ 2.14e+186

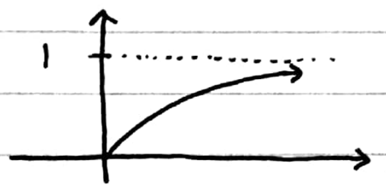
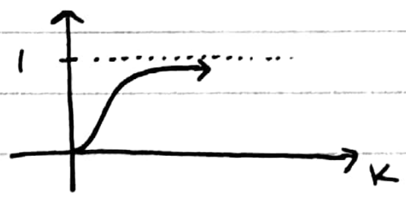
overflow $\rightarrow$ +/- ∞ really big numbers in a computer
underflow $\rightarrow$ 0 get approximated to infinity.

1. Cancel as many terms as you can

$$1 - \frac{365!}{273!\,365^{93}} = 1 - \frac{(365)(364)\dots(273)}{(365)(365)\dots(365)}$$

cancels out a lot for the numerator

$$1 - \frac{n!}{(n-k)!\,n^k}$$ goes up as $k \uparrow$ $P_{n,k}$

2. Stirling's Approximation $\log(n!) = \frac{1}{2}\log(2\pi) + (n+\frac{1}{2})\log(n) - n$
De Moivre first theorized

for any $x > 0$ $x = \exp[\log(x)]$

$$P_{n,k} = 1 - \exp\left[\log\left(\frac{n!}{(n-k)!\,n^k}\right)\right]$$

$$= 1 - \exp[\log(n!) - \log(n-k)! - k\log(n)]$$

3. Gamma Function: generalization of $n!$, n integer, to all positive real numbers - $n! = \Gamma(n-1)$

$$\Gamma(z) = \int_0^\infty e^{z-1} e^{-x}\,dx \quad \Gamma(n) = (n+1)!$$

$$= 1 - \exp[\log(\Gamma(n+1)) - \log(\Gamma(n-k+1)) - k\log(n)]$$

Lecture 4 (cont.)

AMS 131

$\begin{bmatrix} NTS \\ TS \\ TS \end{bmatrix}$ categorical or qualitative $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ $\begin{bmatrix} 73.8 \\ 60.0 \\ 57.9 \end{bmatrix}$ numerical or quantitative

↑ chars string (structural gaps) ↑ discrete int. ↑ decimal float (conceptually continuous)

N distinct elements, $K \geq 2$ groups n_j = group j

$$\sum_{j=1}^{K} n_j = n_i \qquad n \geq 1 \qquad 1 \leq n \leq K \qquad K \geq 2$$

multinomial coefficient

multiple possibilities

$$\binom{n}{n_1, n_2 \ldots n_K} = \frac{n!}{n_1!\, n_2! \ldots n_K!}$$

2016 Presidential Election

random sampling with replacement, n sample

K = 5 outcomes Clinton, Trump, Johnson, Stein, Undecided

$$\sum_{i=1}^{K} p_i = 1$$

IID: each person's outcome independent of others

$P($1st person favors candidate i_1 and 2nd person favors $i_2 \ldots) = p_{i_1} p_{i_2} \ldots p_{i_n}$

$P($sample has x_1 people favoring candidate 1, x_2 choosing $2 \ldots) = p_1^{x_1} p_2^{x_2} \ldots p_K^{x_K}$

$= \square\, p_1^{x_1} \ldots p_K^{x_K}$

$\square$ is the total # of different ways the <u>order</u> of n people in the sample can be listed

$$\square = \binom{n}{x_1, \ldots x_K} = \frac{n!}{x_1!\, x_2! \ldots x_K!}$$

multinomial (prob.) distribution

$$P(X_1 = x_1 \ldots X_n = x_n) = \frac{n!}{x_1! \ldots x_K!} p_1^{x_1} \ldots p_K^{x_K}$$

AMS 131

Lecture 4 (cont.)

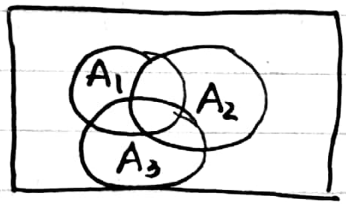
OR with more than two events

$P(A_1 \text{ or } A_2) = P(A_1 \cup A_2)$

$= P(A_1) + P(A_2) - P(A_1 \cap A_2) \leftarrow$ "and"

Kolmogorov's 3RD axiom: if $A_1 \dots A_n$ are disjoint

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$



Ballantine

$P(A_1 \cup A_2 \cup A_3) =$

$P(A_1) + P(A_2) + P(A_3)$

$- [P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)]$

$+ P(A_1 \cap A_2 \cap A_3)$

subtracting the "two at a times" gets rid of the middle, "three at a time"

2 decks of cards with any ordering scheme

- shuffle one deck so all 52! orderings are equally likely
- turn first card of each deck, do they match?

P(at least one match) = ? $n = 52$ cards

A_i = (a match on card i)

we want $P\left(\bigcup_{i=1}^{n} A_i\right) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1}\frac{1}{n!}$

$$\underbrace{\lim_{n \to \infty} \sum_{i=1}^{n} \frac{(-1)^{i+1}}{i!}}_{} = 1 - \frac{1}{e} = 0.63$$

approaches limit quickly - $n = 7$, $= 0.6321$

Metropolis + Ulam

random simulation $\leftrightarrow$ Monte-Carlo method

$\hookrightarrow$ casinos, controlled use of chance

Lecture 4 (cont.) AMS 131

Conditional Probability

$P_K(A)$ A is a set of C (non-weird sets of sample space S)

↑

single argument

he has to define cond. probability $P(A|B) = \begin{cases} \frac{P(A \cap B)}{P(B)} & \text{if } P(B) > 0 \\ \text{undef.} & \text{if } P(B) = 0 \end{cases}$

$P_{def}(A|B)$ or $P_{CJ}(A|B)$ have 2 inputs

all probabilities are conditional on background Assumptions, Information, and Judgment (AIJ)

$P(H) \neq \frac{1}{2}$ coin toss inherently (context)

$P(H \mid \text{fair coin toss}) = \frac{1}{2}$ conditional

This is Bayesian vs. Kolmogorov Frequentist

$P_K(A \mid AIJ)$ vs. $P_K(A)$

1. A, B in C, if $P(B) > 0$ $P(A \cap B) = P(B)P(A|B)$

2. Direct generalization: $A_1 \ldots A_n$

$P(A_1 \cap \ldots \cap A_{n-1}) > 0$ then

$= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \ldots \cap A_{n-1})$

condition based off each previous event

independent? makes it simple, but don't always assume

Partitions: S sample space, events $B_1 \ldots B_K$ such that B_j is disjoint and exhaustive

$\left(\bigcup_{i=1}^{n} B_i = S \right)$

If $(B_1 \ldots B_K)$ is a partition with $P(B_j) > 0$

$P(A) = \sum_{j=1}^{K} P(B_j)P(A|B_j)$ if you know B_j and that A depends on it.

Law of Total Probability weighted by $P(B_j)$ (not all even)

AMS 131 Lecture 4 (cont.)

Weighted mean $\bar{Y}_W = \sum_{i=1}^{n} w_i Y_i$ $\quad 0 \le w_i \le 1 \quad \sum_{i=1}^{n} w_i = 1$

$$P(A) = \sum_{j=1}^{k} P(B_j) P(A|B_j)$$

mixture of $\{P(A|B_j)\}$

<u>mixture of weight $P(B_j)$</u>

weighted average

Assuming all conditional probabilities are defined if c is in C (set), then:

event

$$P(A|C) = \sum_{j=1}^{k} P(B_j|C) P(A|B_j \cap C)$$

A and B are independent iff:

$P(A \cap B) = P(A) \cdot P(B)$ Frequentist

$\left.\begin{array}{l} P(A|B) = P(A) \\ P(B|A) = P(B) \end{array}\right\}$ Bayesian

A, B independent $\xleftrightarrow{\text{iff}}$ information about A doesn't change chances of B, vice versa

Conditional Independence: $\{A_1 \dots A_k\}$ is independent given B if for every subset $\{A_{i_1}, \dots, A_{i_j}\}$ of $\{A_1 \dots A_k\}$ $(j = 2, \dots, k)$

$$P(A_{i_1} \cap \dots \cap A_{i_j} | B) = \prod_{l=1}^{j} P(A_{i_l}|B)$$

$$P(A \text{ and } B | C) = P(A|C) \cdot P(B|C)$$

Machine takes a coin and produces IID so the toss is $P(H) = \theta$, θ can be set $[0,1]$

unknown θ: tosses $Y_1, Y_2 \dots$

<u>1 0 1 1 1 0 0 1 1 1</u> "bits"

H T H H H T T H H H (7 H, 3 T)

Scanned by CamScanner

Lecture 4 (cont.) AMS 131

Can we predict Y_{11}?

A: θ is around $\frac{7}{10}$, so we predict $Y_{11} = H$

THUS Y_{11} depends on $Y_1 \dots Y_{10}$ probabilitistically

Now, $\theta = 0.81$ HHHTHTHHHH (8H, 2T)

if we know θ, the first 10 tosses don't matter, Y_{11} conditionally independent given θ

- given θ, Y_i are independent

The Y_i are unconditionally independent but conditionally independent given θ