Probability GeoComput & ML 03 May 2022









ReCap tot err = comput. + dat.



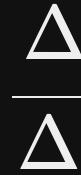
comput error

- machine precision : $\epsilon_{
 m mach}=eta^{1-p}/2$
- representation
- operation

ReCap

question : well posed

solution : well conditioned



condition number : $egin{array}{c} \Delta y/y \ \Delta x/x \end{array}$

ReCap

two cases

- linear system : cond = $\|\mathbf{A}\| \|\mathbf{A}^{-1}\|$
- least square : cond = $\|A\| \|A^+\|$
- projection
- residual

$oldsymbol{A} \| \| oldsymbol{A}^{-1} \| \| oldsymbol{A}^+ \|$

Guided Reading

Why a Hydrology Paper

- very broad : geoscience
- interdisciplinary : nature of geoscience
- math link : eqn 1-3
- personal experience
- class promise

Clouds Physics • outliers -> discovery • Thomas Kuhn : science revolution

Clouds Hydrology • scale • uncertainty

9



Direction

Theory-Guided data science

Direction

Traditional Science

- iteration between data and hypotheses
- knowledge discovery
- knowledge buildup

e d hypotheses

Direction

Data Science

- actionable models
- data under/misrepresentation
- interpretation

Scale

SDM : relating field obs to its environment what's the scale for environment

Scale

Motivating example

coordinate : x $s(z(x), \sigma)$ abundance : N(x) $\lambda(\cdot)$ as a function of its env $\log(\lambda(x))$ z(x)

 $s(z(x),\sigma) = \sum z(x_j)w(x_i,x_j,\sigma)$ $\log(\lambda(x)) = \sum_{i=1}^p eta_i s_i(z_i(x_i),\sigma_i)^{i-1}$

Probability Theory



Basic Concepts

- sample space (S) : the collection of all the outcomes from a random experiment
- ullet event (A) $\subseteq S$
- Prob. function $(P): A \rightarrow \#$

Axioms $ullet P(A) \in [0,1]$ • P(S) = 0• $P(\cup A) = \sum P(A)$

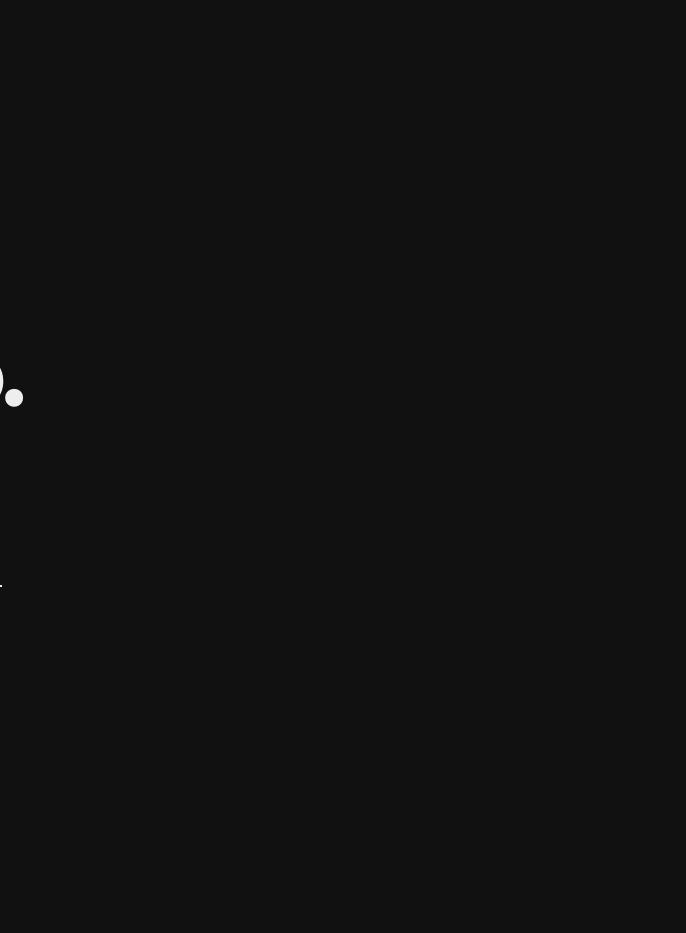


Propositions $P(A^c) = 1 - P(A)$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Propositions Notation $\left[A^{c} ight] = 1 - \left[A ight]$ [A+B] = [A] + [B] - [A,B]



Cond. Prob. $[A|B] = \frac{[A,B]}{[B]}$



Cond. Prob.

- Example:
- [rain,Sat] = [rain,Sun] = 0.5
- [rain, two conted days] = 0.6
 - [rain, weekend] ?

= 0.5 = 0.6

Cond. Prob.

- [Sat + Sun] = [Sat] + [Sun] [Sat, Sun]
 - [Sat, Sun] = [Sun | Sat] [Sat] = 0.3
 - [Sat + Sun] = 1 [Sat, Sun] = 0.7

• [Sat, Sun] [sat, Sun] [sat, Sun]

Independence [A,B] = [A][B]



Independence Example : [rain, Sat] = [rain, Sun] = 0.5Sat 🔟 Sun [rain, weekend]?



Independence [Sat, Sun] = [Sat] + [Sun] = 0.25[Sat + Sun] = 0.75

Law of Total Probability $[B] = \sum [B|A][A]$

Law of Total Probability Kidney Stone Treatment

	A	B
S	81/87=0.93	234/27
L	192/263=0.73	55/80=
	273/350=0.78	289/3

[E|A] = [E|A,S][S|A] + [E|A,L][L|A]

70=0.87

=0.69

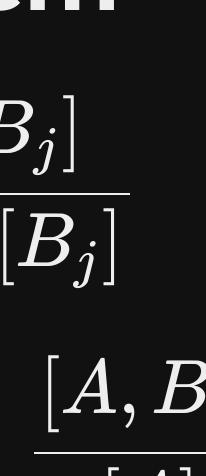
50=0.83

Law of Total Probability Bonus

[E|A] = [[S+L,E]|A]= [[S, E] + [L, E]|A]= [[S,E]|A] + [[L,E]|A] $\left[\left[S,E,A ight] +\left[L,E,A ight] ight]$ [A]= [E|S,A][S,A] + [E|L,A][L,A]



Bayes Theorem $[B_j|A] = rac{[A|B_j][B_j]}{\sum [A|B_j][B_j]}$ $[B_j|A] = rac{[A|B_j][B_j]}{[A]} = rac{[A,B_j]}{[A]}$

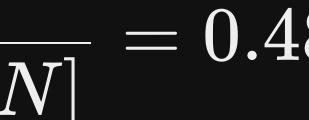


Bayes Theorem

- Example :
- 1% pop have cancer: [Y] = 0.01
- 80% test + if cancer : [+|Y] = 0.8
- 9.6% test + if no cancer : [+|N] = 0.096

[Y | +] = ?

 $[Y|+] = rac{[+|Y][Y]}{[+|Y][Y] + [+|N][N]} = 0.48$



Random Variable RV : real valued function mapped onto the sample

RV : real valued function mapped space

Random Variable Example

- flip a coin twice, denote X as the # of heads
 - X(TT)=0, X(TH)=X(HT)=1, X(HH)=2
 - [X=0]=1/4, [X=1]=1/2, [X=2]=1/4
 - Prob. Distr. of X

e # of heads (HH)=2 =2]=1/4

Random Variable pmf $[x_k] = [X = x_k], k = 1, 2, 3...$

Expectation Value

 $E(X) = \sum x_k[x_k]$

Expectation Value

coin game



 $iggl\{ egin{array}{c} [H,U] = x \ [H,I] = y \end{array}
ight.$

 $A \Rightarrow E(U) = 3xy + (1-x)(1-y) - 2(x(1-y) + y(1-x))$



Prob. Distr. Functions Binomial

 $[k;n,p]=inom{n}{k}p^k(1-p)^{(n-k)}$

Prob. Distr. Functions Poisson $[k;\lambda] = rac{\lambda^k e^{-\lambda}}{k!}$

Prob. Distr. FunctionsNormal $[x;\mu,\sigma]=(\sigma\sqrt{2\pi})^{-1}\exp(-(x-\mu)^2/2\sigma^2),\ x\in\mathbb{R}$

Acknowledgement

Thanks for Your Attention

References

- N. Cressie and C. K. Wikel, Statistics for Spatio-Temporal Data, 2011
- N. Cressie, Statistics for Spatial Data, 1991
- P. Olofsson and M. Andersson, Probability, Statistics, and Stochastic Processes, 2012
- R. Larsen and M. Marx, An introduction to mathematical statistics and its applications, 2018
- A. Agresti, Foundations of Linear and Generalized Linear Models, 2015
- C. McMulloch, et.al., Generalised, Linear and Mixed Models, 2008
- https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020WR028091
- https://ieeexplore.ieee.org/document/7959606
- https://link.springer.com/article/10.1007/s10980-016-0380-z
- https://en.wikipedia.org/wiki/Simpson%27s_paradox

