Probability Theory

GeoComput & ML

03 May 2022

ReCap

tot err = comput. + dat.

ReCap

comput error

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

ReCap

question : well posed

solution : well conditioned

condition number : $\left| \dfrac{\Delta y / y}{\Delta x / x} \right|$

ReCap

two cases

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

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

Clouds

Direction

Theory-Guided data science

Direction

Traditional Science

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

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

abundance : $N(x)$

$\lambda(\cdot)$ as a function of its env $z(x)$

$$s(z(x), \sigma) = \sum z(x_j) w(x_i, x_j$$

$$\log(\lambda(x)) = \sum_{i=1}^{p} \beta_i s_i(z_i(x_i), \sigma_i$$

Probability Theory

Basic Concepts

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

Axioms

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

Propositions

$$P(A^c) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Propositions

Notation

$$[A^c] = 1 - [A]$$

$$[A + B] = [A] + [B] - [A, B]$$

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] ?

Cond. Prob.

[Sat + Sun] = [Sat] + [Sun] - [Sat, Sun]

[Sat, Sun] = [Sun|Sat] [Sat] = 0.3

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

Independence

$$[A, B] = [A][B]$$

Independence

Example :

[rain, Sat] = [rain, Sun] = 0.5

Sat $\perp\!\!\!\perp$ 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	234/270=0.87
L	192/263=0.73	55/80=0.69
	273/350=0.78	289/350=0.83

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

Law of Total Probability

Bonus

$$
\begin{aligned}
[E|A] &= [[S+L,E]|A] \\
&= [[S,E]+[L,E]|A] \\
&= [[S,E]|A]+[[L,E]|A] \\
&= \frac{[S,E,A]+[L,E,A]}{[A]} \\
&= [E|S,A][S,A]+[E|L,A][L,A]
\end{aligned}
$$

Bayes Theorem

$$[B_j|A] = \frac{[A|B_j][B_j]}{\sum[A|B_j][B_j]}$$

$$[B_j|A] = \frac{[A|B_j][B_j]}{[A]} = \frac{[A, B_j]}{[A]}$$

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|+] = ?

$$[Y|+] = \frac{[+|Y][Y]}{[+|Y][Y] + [+|N][N]} = 0.48$$

Random Variable

RV : real valued function mapped onto the sample 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

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

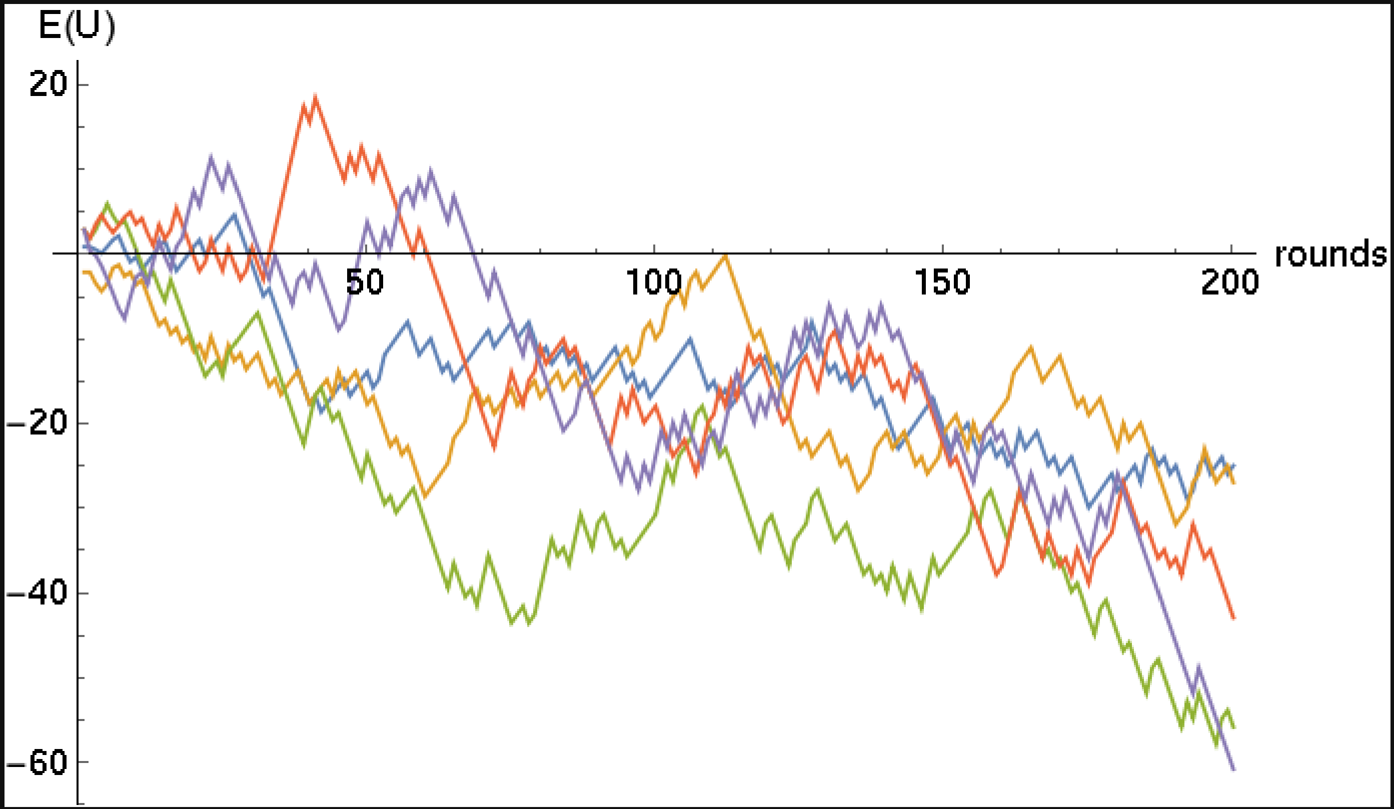
	H	**T**
H	3	-2
T	-2	1

$$\begin{cases} [H,U] = x \\ [H,I] = y \end{cases}$$

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

Expectation Value

Simulation : $[U, H] \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$



Prob. Distr. Functions

Binomial

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

Prob. Distr. Functions

Poisson

$$[k; \lambda] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Prob. Distr. Functions

Normal

$$[x;\mu,\sigma] = (\sigma\sqrt{2\pi})^{-1}\exp(-(x-\mu)^2/2\sigma^2),\ x \in$$

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