METO 630 Class Notes (Eugenia Kalnay)

Review of Probability, Wilks, Chapter 2

Events: elementary and compound, E
Sample space: space of all possible events, S
MECE: Mutually exclusive and collectively exhausting events
Probability Axioms:

$P(A) \geq 0;$

$P(S) = 1;$

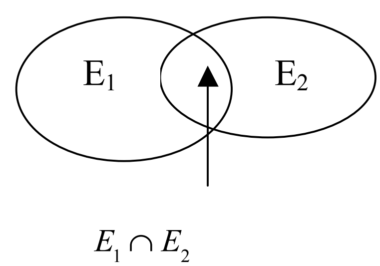
If $(E_1 \cap E_2) = 0$, i.e., if E_1 and E_1 exclusive, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

Probability ~ Frequency $P(E) = \lim_{n \to \infty} \frac{\# E = yes}{\text{total n}}$

Venn diagrams

If $E_2 \subseteq E_1$, then $P(E_1) \geq P(E_2)$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

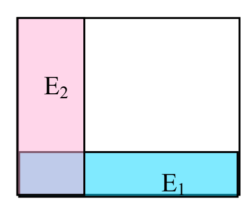


Recall threat score: $TS = \frac{P(F = yes \cap Ob = yes)}{P(F = yes \cup Ob = yes)}$

Conditional Probability: "probability of E_1 given that E_2 has happened"

$$P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

Independent events: $P(E_1 \cap E_2) = P(E_1) P(E_2)$



This means that

$$P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = P(E_1)$$

i.e., the probability of E_1 happening is independent of whether E_2 happened (e.g., the probability of a summer storm is independent from the phases of the moon).

Exercise: From the Penn State station data for January 1980, compute the probability of precipitation, of T>32F, conditional probability of pp if T>32F, and conditional probability of pp tomorrow if it is raining today.

Exercise: Prove graphically the DeMorgan Laws:

$$P\{(A \cup B)^C\} = P\{A^C \cap B^C\}; P\{(A \cap B)^C\} = P\{A^C \cup B^C\}$$

Total probability:

$$P(A) = \sum_{i=1}^{I} P(A \cap E_i) = \sum_{i=1}^{I} P(A \mid E_i) P(E_i)$$ where E_i are MECE.

Bayes Theorem: It "inverts" the probability

$$P(E_i \mid A) = \frac{P(E_i \cap A)}{P(A)} = \frac{P(A \mid E_i)P(E_i)}{P(A)} = \frac{P(A \mid E_i)P(E_i)}{\sum_{j=1}^{I} P(A \mid E_j)P(E_j)}$$

Combines prior information with new information

Example of Bayesian reasoning:
Relationship between pp over SE US and El Niño

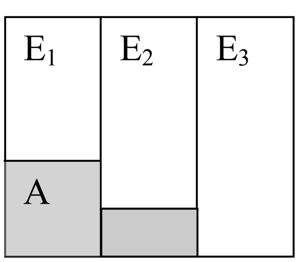
Precip. Events: E_1(above), E_2(normal), E_3(below) are MECE. A is El Niño

Prior information (from past statistics):

$$P(E_1) = P(E_2) = P(E_3) = 33\%$$
$$P(A \mid E_1) = 40\%;\ P(A \mid E_2) = 20\%;\ P(A \mid E_3) = 0\%;$$

Total probability of A:

$$P(A) = P(A \mid E_1)P(E_1) + P(A \mid E_2)P(E_2) + P(A \mid E_3)P(E_3) =$$
$$P(A) = 40\%\ 33\% + 20\%\ 33\% + 0\%\ 33\% = 20\%$$



Bayes, new information: El Niño is happening!!

What is the probability of above normal precipitation?
Note the clear interpretation from the figure: once you know A is true, the prob. of E_1 is 2/3.

$$P(E_1 \mid A) = \frac{P(A \mid E_1)}{P(A)} = \frac{40\%\ 33\%}{20\%} = 66\%!$$

Example of Bayesian use in variational data assimilation:

Prior knowledge (measurement or forecast) T_1 of the true value T

New measurement: T_2.

$$P(T \mid T_2) = \frac{P(T_2 \mid T) P_{prior,givenT_1}(T)}{P(T_2)} = \frac{\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(T_2 - T)^2}{2\sigma_2^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(T - T_1)^2}{2\sigma_1^2}}}{e^{-\frac{(T_2 - \bar{T})^2}{2\sigma_2^2}}}$$

Note that the total probability of a measurement T_2 given a climatological average $\bar{T}$ is independent of *T*.

We choose as our best estimate of the true temperature *T* the value that maximizes (over *T*) the probability $P(T \mid T_2)$. Since the logarithm is monotonic, it is equivalent to maximizing (over *T*) the $\log P(T \mid T_2)$,

$$\log P(T \mid T_2) = const - \frac{(T_2 - T)^2}{2\sigma_2^2} - \frac{(T - T_1)^2}{2\sigma_1^2}$$

or minimize (over *T*) the cost function used in 3D-Var:

$$J = \frac{(T - T_2)^2}{2\sigma_2^2} + \frac{(T - T_1)^2}{2\sigma_1^2}.$$