Review of Probability
Wilks, Chapter 2

Ji-Sun Kang and Eugenia Kalnay

AOSC630
Spring, 2008, updated in 2017
Definition

- **Event**: set, class, or group of possible uncertain outcomes
  - *A compound event* can be decomposed into two or more (sub) events
  - *An elementary event* cannot be decomposed
- **Sample space (event space), S**
  - The set of all possible elementary events
- **MECE (Mutually Exclusive & Collectively Exhaustive)**
  - *Mutually Exclusive*: no more than one of the events can occur
  - *Collectively Exhaustive*: at least one of the events will occur
  - → *A set of MECE events completely fills a sample space*
Probability Axioms

- $P(A) \geq 0$
- $P(S) = 1$
- If $(E_1 \cap E_2) = 0$, i.e., if $E_1$ and $E_2$ exclusive, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$
Probability

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

- **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)**

\[
\frac{P(F = \text{yes} \cap \text{Ob} = \text{yes})}{P(F = \text{yes} \cup \text{Ob} = \text{yes})}
\]

**Venn Diagrams**
Conditional Probability

- **Probability of $E_1$ given that $E_2$ has happened**

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

\[ P(E_1 \cap E_2) = P(E_1 | E_2) P(E_2) \quad ; \text{Multiplicative Law} \]

- **Independent Event**
  - The occurrence or nonoccurrence of one does not affect the probability of the other

\[ P(E_1 \cap E_2) = P(E_1) P(E_2) \]

i.e. \[ P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = P(E_1) \]
Exercise

- From the Penn State station data for Jan. 1980, compute the probability of precipitation, of $T>32F$, conditional probability of precipitation if $T>32F$, and conditional probability of precipitation tomorrow if it is raining today.

- Prove graphically the DeMorgan Laws:
  
  $P\left\{ (A \cup B)^c \right\} = P\left\{ A^c \cap B^c \right\}; \quad P\left\{ (A \cap B)^c \right\} = P\left\{ A^c \cup B^c \right\}$
Total Probability

- MECE events, \( \{E_i\}, \; i=1, \ldots, I \)

\[
P(A) = \sum_{i=1}^{I} P(A \cap E_i) = \sum_{i=1}^{I} P(A|E_i)P(E_i)
\]
Bayes’ Theorem

- **Bayes’ theorem is used to “invert” conditional probabilities**
  - If $P(E_1|E_2)$ is known, Bayes’ Theorem may be used to compute $P(E_2|E_1)$.

\[
P(E_i|A) = \frac{P(E_i \cap A)}{P(A)} = \frac{P(A|E_i)P(E_i)}{P(A)} = \frac{P(A|E_i)P(E_i)}{\sum_{j=1}^{I} P(A|E_i)P(E_j)}
\]

- Combines prior information with new information
Example of Bayesian Reasoning

- Relationship between precipitation over SE US and El Nino
  - Precipitation Events: $E_1$(above), $E_2$(normal), $E_3$(below) are MECE
  - El Nino Event: $A$
  - Prior information (from past statistics)
    - $P(E_1)=P(E_2)=P(E_3)=33\%$
    - $P(A|E_1)=40\%$; $P(A|E_2)=20\%$; $P(A|E_3)=0\%$

\[
P(A|E_1) = \frac{P(A \cap E_1)}{P(E_1)} = \frac{0.40}{0.33} \approx 0.40
\]

\[
P(A|E_2) = \frac{P(A \cap E_2)}{P(E_2)} = \frac{0.20}{0.33} \approx 0.20
\]
Example of Bayesian Reasoning

- **Total probability of A**
  \[
P(A) = \sum_{i=1}^{3} P(A \mid E_i)P(E_i)
  \]
  \[
  = P(A \mid E_1)P(E_1) + P(A \mid E_2)P(E_2) + P(A \mid E_3)P(E_3)
  \]
  \[
  = 0.4 \times 0.33 + 0.2 \times 0.33 + 0 \times 0.33 = 0.20
  \]

- **NEW information**: El Nino is happening!
  Probability of above normal precipitation?

  \[
P(E_1 \mid A) = \frac{P(A \cap E_1)}{P(A)}
  \]
  \[
  = \frac{0.4 \times 0.33}{0.2} = 0.66
  \]
Probability Density Function when we have two observations, \( T_1 \) and \( T_2 \)

With Gaussian errors, the best estimate of \( T \) is the weighted average of \( T_1 \) and \( T_2 \):

The weights are proportional to the error variance of the other obs.:

\[
T = \frac{\sigma_2^2 T_1 + \sigma_1^2 T_2}{\sigma_1^2 + \sigma_2^2}
\]

Precision = inverse of \( \sigma^2 \)

Precision of \( T \) = Precision of \( T_1 \) + Precision of \( T_2 \):

\[
\frac{1}{\sigma_T^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}
\]
If we have a forecast $T_f$ (prior info.) and then an observation $T_o$

The optimal analysis $T_a$ is the best estimate of the truth (it has minimum errors):

$$T_a = \frac{\sigma_o^2 T_f + \sigma_f^2 T_o}{\sigma_f^2 + \sigma_o^2}$$

The analysis $T_a$ is more accurate than both the forecast $T_f$ and the observation $T_o$!

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_f^2} + \frac{1}{\sigma_o^2}$$

These are the formulas used in data assimilation!
Now let’s use a Bayesian approach for Data Assimilation

Bayes theorem

\[ P(T \mid T_o) = \frac{P(T \mid T_f)P(T_o \mid T)}{P(T_o)} \]

“The posterior probability of the true temperature \( T \) given the prior information \( T_f \), and after receiving the new observation \( T_o \), is given by the prior probability of \( T \) (based on the forecast \( T_f \)) multiplied by the likelihood of \( T \) given the observation \( T_o \), normalized by the total probability of obtaining a measurement \( T_o \).”

The likelihood of \( T \) given the observation \( T_o \) is the same as the probability of observing \( T_o \) given a true temperature \( T \) (Edwards, 1984). This formula can be briefly read as:

“\[ \text{posterior} = \text{prior} \cdot \text{likelihood} / \text{normalization} \]"
The Bayesian approach for Data Assimilation is very general (not just Gaussians)

If we assume Gaussianity, the Bayes theorem leads to the Variational approach:

\[
P(T \mid T_o) = \frac{P(T \mid T_f)P(T_o \mid T)}{P(T_o)} =
\]

\[
e^{-\frac{(T-T_f)^2}{2\sigma^2_f}} \frac{\sqrt{2\pi\sigma^2_o}}{\sqrt{2\pi\sigma^2_f}} \frac{e^{-\frac{(T_o-T)^2}{2\sigma^2_o}}}{e^{-\frac{(T_o-T_{cli})^2}{2\sigma^2_o}}} =
\]

\[
const * e^{-\frac{(T-T_f)^2}{2\sigma^2_f} + \frac{(T_o-T)^2}{2\sigma^2_o}}
\]

Since we want to maximize the probability of \(T\), and \(T_{cli}\), the climatological temperature probability distribution does not depend on \(T\), the maximization can be written as the minimization of the exponent:
The Bayesian approach for Data Assimilation is very general (not just Gaussians)

From Bayes theorem, we minimize the exponent, a cost function $J$ that measures the squared distance between the optimal temperature $T$ that we are seeking and the prior forecast, and with the new observation, both normalized by their error variances:

$$J = \frac{(T - T_f)^2}{2\sigma_f^2} + \frac{(T_o - T)^2}{2\sigma_o^2}$$

Although the variational formulation looks very different from the classical formulation shown before, for Gaussian errors, both give the same solution. However, the Bayesian approach can be used with any probability distributions and allows the implementation of efficient “particle filters” (e.g., Penny and Miyoshi, 2016, Poterjoy, 2016).