Motivation

- Modeling uncertainty
  1) inherent randomness in world
  2) gross statistical description of complex deterministic world
  3) guardian of commonsense reasoning

Review of probability

- Discrete random variables \( X \) (capitalized)
  Domain of possible values \( \{X_1, X_2, \ldots, X_n\} \) (lower case)

Ex: month \( M \), \( \{M_1=\text{JAN}, M_2=\text{FEB}, \ldots, M_{12}=\text{DEC}\} \)

- "Unconditional" or "prior" probability: \( P(X=x) \)
  Basic axioms:
  (i) \( P(X=x) \geq 0 \) "probability the event \( X=x \) is true"
  (ii) \( \sum_{i=1}^{n} P(X=x_i) = 1 \)
  (iii) \( P(X=x_i \text{ or } X=x_j) = P(X=x_i) + P(X=x_j) \text{ if } x_i \neq x_j \)
  Probs add for union of mutually exclusive events.

- "Conditional" or "posterior" probability
  \( P(X=x_i | Y=y_j) \) = probability that \( X=x_i \) given that \( Y=y_j \)

In general: \( P(X=x_i | Y=y_j) \neq P(X_i) \)

Ex: conditional dependence
  weather \( W \) \( \{w_1=\text{sunny}, w_2=\text{rainy}\} \)
  \( P(W=\text{sunny}) = 0.9 \)
  \( P(W=\text{sunny} | M=\text{jan}) = 0.7 \) depending on condition, probability
  \( P(W=\text{sunny} | M=\text{aug}) = 0.95 \)
  Can change either way...

Ex: conditional independence
  day of week \( D \) \( \{d_1=\text{mon}, \ldots, d_7=\text{sun}\} \)
  \( P(W=\text{rain} | D=\text{tues}) = P(W=\text{rain}) \)

Also true:
(i) \( P(X=x_i | Y=y_j) \geq 0 \)
(ii) \( \sum_{i} P(X=x_i | Y=y_j) = 1 \) Note: sum over \( i \), not \( j \)
• Joint probability
  \( P(X=x_i, Y=y_j) = \text{prob. that } X=x_i \text{ and } Y=y_j \)

• Product rule: from conditional to joint
  For all \( i, j \):
  \( P(X=x_i, Y=y_j) = P(X=x_i \mid Y=y_j) P(Y=y_j) \)
  Also:
  \( P(X=x_i, Y=y_j) = P(Y=y_j \mid X=x_i) P(X=x_i) \)

• Generalized product rule: create composite event from simpler events
  \( P(A=a_i, B=b_j, C=c_k, D=d_l, \ldots) \)
  \( = P(A=a_i) P(B=b_j \mid A=a_i) P(C=c_k \mid A=a_i, B=b_j) P(D=d_l \mid A=a_i, B=b_j, C=c_k) \)

• Easier to assess conditional probabilities (RHS) than joint probs (LHS)
  Ex: \( A = \text{wake up on time} \)
    \( B = \text{eat breakfast} \)
    \( C = \text{hit traffic} \)
    \( D = \text{arrive on time UCSD} \)

• Marginalization: from joint to marginal distribution
  \( P(X=x_i) = \sum_j P(X=x_i, Y=y_j) \)
  \( P(X=x_i, Y=y_j) = \sum_k P(X=x_i, Y=y_j, Z=z_k) \)
  In this context, probabilities on LHS (over smaller subsets of random variables) are known as \textit{marginal} probabilities.

• Shorthand notation
  (i) implied universality: \( P(X,Y) = P(X \mid Y) P(Y) = P(Y \mid X) P(X) \)
  implied that equality holds for all \( X=x_i \) and \( Y=y_j \)
  (ii) implied assignment: \( P(x, y, z) = P(X=x, Y=y, Z=z) \)
  ok to omit assignment when context is unambiguous
  Ex: generalized product rule
  \( P(a, b, c, d, \ldots) = P(a) P(b \mid a) P(c \mid a, b) \ldots \)
Bayes rule relates conditional probs to other conditional probs.

From "inplied unversity equation,
\[ P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \]
"if you see the effect, you can infer the cause."

Intuitively, if \( X \) is a "cause" and \( Y \) is an "effect," we can use Bayes rule to infer \( X \) from \( Y \).

Ex: cancer diagnosis

Given: 1% population has cancer
Test has 10% false negative rate.
Test has 20% false positive rate.

• You test positive. Do you have cancer?

• Random variables
  DIAGNOSIS \( \in \{ \text{cancer, healthy}\} \)
  TEST \( \in \{ \text{pos, neg}\} \)

• Probabilities:
  \[ P(\text{cancer}) = 0.01 \]
  \[ P(\text{pos} \mid \text{cancer}) = 0.9 \]
  \[ P(\text{pos} \mid \text{healthy}) = 0.2 \]

• Marginalization
  \[ P(\text{pos}) = \sum_{\text{DIAGNOSIS}} P(\text{pos}, \text{DIAGNOSIS}) = \sum_{\text{DIAGNOSIS}} P(\text{pos} \mid \text{DIAGNOSIS})P(\text{DIAGNOSIS}) \]
  \[ = P(\text{pos} \mid \text{cancer})P(\text{cancer}) + P(\text{pos} \mid \text{healthy})P(\text{healthy}) \]
  \[ = (0.9)(0.01) + (0.2)(1 - 0.01) = 0.207 \]

• Bayes rule
  \[ P(\text{cancer} \mid \text{pos}) = \frac{P(\text{pos} \mid \text{cancer})P(\text{cancer})}{P(\text{pos})} = \frac{(0.9)(0.01)}{0.207} \]
  \[ = 0.043 = \frac{4.3\%}{\text{Before test}} \]
  \[ = 0.01 \text{ or 1%} \]
  \[ = P(\text{cancer} \mid \text{pos}) = 4.3\% \]
  \[ = P(\text{cancer} \mid \text{pos}) \ll P(\text{pos} \mid \text{cancer}) = 90\% \]

All terms in Bayes rule are important.
Conditioning on background evidence often useful to reason in context of background knowledge. Consider events $X$ and $Y$ and background evidence $E$.

(i) conditionalized version of product rule

$$P(X, Y | E) = \frac{P(X, Y, E)}{P(E)} = \frac{P(X, Y, E)}{P(Y, E)} \cdot \frac{P(Y, E)}{P(E)} = \frac{P(X | Y, E)}{P(Y | E)} \cdot \frac{P(Y | E)}{P(E)}$$

"this is useful"

Also: $$P(X, Y | E) = P(Y | X, E) \cdot P(X | E)$$

(ii) conditionalized version of Bayes rule

From above: $$P(X | Y, E) \cdot P(Y | E) = P(Y | X, E) \cdot P(X | E)$$

$$\Rightarrow P(X | Y, E) = \frac{P(Y | X, E) \cdot P(X | E)}{P(Y | E)}$$

• Conditional independence statements

  The following three statements are equivalent:

  (i) $$P(X, Y | Z) = P(X | Z) \cdot P(Y | Z)$$

  (ii) $$P(X | Y, Z) = P(X | Z)$$

  (iii) $$P(Y | X, Z) = P(Y | Z)$$

  Proof: HW

  If any one is true, the other two are true.
Kullback–Leibler divergence (KL)

How to measure the difference between two probability distributions?

Let: 
\[ p_i = P(X=x_i | E) \]
\[ q_i = P(X=x_i | E') \]

Conditioned on different evidence \( E \neq E' \)

Define: 
\[ KL(p, q) = \sum_i p_i \log \left( \frac{p_i}{q_i} \right) \]

Properties of KL "distance" (not really a distance since it isn't symmetric)

(i) \( KL(p, q) \geq 0 \) with equality if \( p_i = q_i \) for all \( i \)
(ii) \( KL(p, q) \neq KL(q, p) \) in general