Terms and concepts from last time:

- Observation, target
- Attribute, feature
- Regression, classification
- Supervised, unsupervised, reinforcement
- Model, model family, parameter, hyperparameter
- Error function, overfitting, loss function

Let S be a sample space. A *probability function* $P : \mathscr{P}(S) \to [0, 1]$ fulfills the axioms of probability:

1. For all $X \in \mathscr{P}(S)$, $P(X) \geq 0$.
2. $P(S) = 1$
3. For disjoint sets $X, Y \in \mathscr{P}$, $P(X \cup Y) = P(X) + P(Y)$.

$$P(\{\heartsuit J, \heartsuit Q, \heartsuit K, \diamondsuit J, \ldots \clubsuit K\}) = \frac{12}{52} \approx .23$$

Consider the events

- A, the card is red. $P(A) = .5$
- B, the card is a diamond. $P(B) = .25$
- C, the card is a 4. $P(C) \approx .077$
- D, the card is $\diamondsuit 4$. $P(D) \approx .019$

Theorem 1 (Bayes's theorem)

If X and Y are events, $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$.

Proof.

$$\begin{aligned} P(Y|X) &= \frac{P(X \cap Y)}{P(X)} \\ P(X \cap Y) &= P(Y|X)P(X) \\ P(X|Y) &= \frac{P(X \cap Y)}{P(Y)} \\ &= \frac{P(Y|X)P(X)}{P(Y)} \end{aligned}$$

□

A *probability mass function* of a random variable X is a function from values X can take on to the probability that X will take on that value. $p_X : \mathbb{R} \to [0,1]$.

$$\begin{aligned} p_X(x) &= P(X = x) \\ &= P(A_x) \\ &= \textstyle\sum_{s \in S \ | X(s) = x} P(\{s\}) \end{aligned}$$