Probability Problems

> What are the general tasks we expect to solve with probabilistic programs?

Background

- **Conditional Probability**
$$P(A, B) = P(B|A)P(A).$$
- **Bayes Theorem**
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$
- **For maximization tasks**
$$P(B|A) \propto P(A|B)P(B).$$
- **Marginal**
$$P(A) = \sum_b P(A, b).$$
- In $P(B|A) \propto P(A|B)P(B)$, if the **posterior** $P(B|A)$ and the **prior** $P(B)$ follow distributions of the same family, $P(B)$ is a **conjugate prior** for the **likelihood** $P(A|B)$.
- **Density Estimation:** Estimate a joint probability distribution from a set of observations; Select a probability distribution function and the parameters that best explains the distributions of the observations.

MLE: Maximum Likelihood Estimation

> Given a probability **distribution** d and a set of **observations** X, find the distribution **parameters** θ that maximize the **likelihood** (*i.e.* the probability of those observations) for that distribution.
>
> **Overfits the data:** high variance of the parameter estimate; sensitive to random variations in the data. Regularization with $P(\theta)$ leads to **MAP**.

Given d, X, find
$$\hat{\theta}_{\mathsf{MLE}}(d, X) = \arg_\theta \max P_d(X|\theta).$$

MAP: Maximum A-Priori

> Given a probability **distribution** d and a set of **observations** X, find the distribution **parameters** θ that best explain those observations.

Given d, X, find
$$\hat{\theta}_{\mathsf{MAP}}(d, X) = \arg_\theta \max P(\theta|X).$$

Using $P(A|B) \propto P(B|A)P(A)$,
$$\hat{\theta}_{\mathsf{MAP}}(d, X) = \arg_\theta \max P_d(X|\theta)P(\theta)$$

Variants: - **Viterbi algorithm:** Find the most likely sequence of hidden states (on HMMs) that results in a sequence of observed events. - **MPE: Most Probable Explanation** and **Max-sum, Max-product algorithms:** Calculates the marginal distribution for each unobserved node, conditional on any observed nodes; Defines the most likely assignment to all the random variables that is consistent with the given evidence.