Probability, Estimation, and Sampling

Discrete Probability Distributions

- Event space $A = \{1, 2, \ldots, n\}$
- Distribution $P : A \rightarrow [0,1], \sum_{a \in A} P(a) = 1$
- Intuitively, $P(a)$ is the chance that a happens
- Networks output distributions

Conditional Probability

- Chance that a happens given θ: $P(a|\theta)$

- E.g., roll two 6-sided dice and call the results a and b

$$P(a+b=12)=\frac{1}{36} \qquad P(a+b=12|a=6)=\frac{1}{6}$$

- Model parameters

Likelihood

- A function of θ describing the probability of observing data x given θ

$$L(\theta) = L(\theta; x) = P(x|\theta)$$

Sampling

$a \sim P$: Generate events a according to distribution P

Sampling bias: Observed data does not follow the distribution

Expected Value

$$\mathrm{E}_{a\sim \mathrm{P}}[f(a)] = \sum_{a} P(a) f(a) \approx \frac{1}{N} \sum_{a\sim P} f(a)$$

$$\mathrm{E}_{a\sim \mathrm{P}}[c f(a)] = c\,\mathrm{E}_{a\sim \mathrm{P}}[f(a)]$$

$$\mathrm{E}_{a\sim \mathrm{P}}[f(a)+g(a)] = \mathrm{E}_{a\sim \mathrm{P}}[f(a)]+\mathrm{E}_{a\sim \mathrm{P}}[g(a)]$$

$$\mathrm{E}_{a\sim \mathrm{P}}[f(a) g(a)] \neq \mathrm{E}_{a\sim \mathrm{P}}[f(a)]\mathrm{E}_{a\sim \mathrm{P}}[g(a)]$$