Lecture 06: Probability Basics

Probability Distributions

- Ω is the sample space (i.e., the set of elements to be sampled)
- A is a probability distribution with sample space Ω
- $A(i)$ represents the probability $\Pr[A = i]$, i.e. the probability of sampling $i \in \Omega$ according to the distribution A

Statistical Distance

Suppose Ω is a finite size sample space

Definition (Statistical Distance)

$$\mathrm{SD}\left(A, B\right) := \frac{1}{2} \sum_{i \in \Omega} \left|A(i) - B(i)\right|$$

Intuition: $\mathrm{SD}\left(A, B\right)$ represents (half) the area between the curves A and B. If two curves have small region between them then the two curves look similar. So, $\mathrm{SD}\left(A, B\right)$ being small implies that the probability distributions A and B are similar.

Some Properties

- Note: If $A(i) = B(i)$ then the i-th summand in the statistical distance definition has no contribution
- Let Ω_A be the set of all i such that $A(i) \geqslant B(i)$. Formally written as: $\Omega_A = \{i \colon i \in \Omega, A(i) \geqslant B(i)\}$
- Let Ω_B be the set of all i such that $A(i) < B(i)$. Formally written as: $\Omega_B = \{i \colon i \in \Omega, A(i) < B(i)\}$
- Note that: Ω_A and Ω_B partition Ω
- Think:

Claim

$$\sum_{i \in \Omega_A} A(i) - B(i) = \sum_{i \in \Omega_B} B(i) - A(i) = \mathrm{SD}\left(A, B\right)$$

Alternate Equivalent Definition

- An event E is a subset of Ω
- The probability of E according to probability distribution A is represented by $A(E)$ and is equal to $\sum_{i \in E} A(i)$

Definition (Statistical Distance)

$$\max_{E \subseteq \Omega} A(E) - B(E)$$

Equivalence

Claim

$$\frac{1}{2}\sum_{i\in\Omega}|A(i)-B(i)| = \max_{E\subseteq\Omega} A(E)-B(E)$$

- Let E^* be an event that achieves the maximum value$\max_{E\subseteq\Omega} A(E)-B(E)$
- First observation: E^* cannot contain $i\in\Omega_B$. Proof: Suppose $i\in\Omega_B$ and $i\in E^*$. Note that $A(i)-B(i)$ is negative. Let E' be the event $E^*\setminus\{i\}$. Note that $A(E')-B(E')$ is greater than $A(E^*)-B(E^*)$. This contradicts the maximality of $A(E^*)-B(E^*)$.
- Think: Why should E^* contain all $i\in\Omega$ such that $A(i)>B(i)$?
- Without loss of generality, we can assume that $E^*=\Omega_A$
- For this choice, it is easy to see that both definitions are equal

Triangle Inequality

Claim (Triangle Inequality)

$$\mathrm{SD}\left(A, B\right) \leqslant \mathrm{SD}\left(A, C\right) + \mathrm{SD}\left(C, B\right)$$

- Follows from the following manipulation

$$\begin{aligned}
\mathrm{SD}\left(A, B\right) &= \frac{1}{2}\sum_{i\in\Omega} |A(i) - B(i)| \\
&= \frac{1}{2}\sum_{i\in\Omega} |A(i) - C(i) + C(i) - B(i)| \\
&\leqslant \frac{1}{2}\sum_{i\in\Omega} |A(i) - C(i)| + |C(i) - B(i)| \\
&= \mathrm{SD}\left(A, C\right) + \mathrm{SD}\left(C, B\right)
\end{aligned}$$

- Think: Equality holds if and only if $C(i)$ is between $A(i)$ and $B(i)$, for all $i \in \Omega$

Distribution of a Function Output

- Let $f\colon \Omega \to \Omega'$ be a function
- The output distribution $f(x)$ where x is sampled according to the distribution A is represented by $f(A)$
- The probability of $f(A)$ outputting y is represented by $(f(A))(y)$
- Suppose $y \in \Omega'$
- Let $f^{-1}(y)$ be the set of all $x \in \Omega$ such that $f(x) = y$
- The probability of outputting y is given by the probability of sampling an element in $f^{-1}(y)$ according to the distribution A, i.e., $A(f^{-1}(y))$ or equivalently $\sum_{x \in f^{-1}(y)} A(x)$
- Note that the sets $f^{-1}(y)$, for $y \in \Omega'$, partition Ω

Data-processing Inequality

Claim (Data-processing Inequality)

For any function f, the following holds:

$$\mathrm{SD}\left(f(A), f(B)\right) \leqslant \mathrm{SD}\left(A, B\right)$$

- Consider the following manipulation:

$$\begin{aligned}
\mathrm{SD}\left(f(A), f(B)\right) &= \frac{1}{2} \sum_{y \in \Omega'} \left|(f(A))(y) - (f(B))(y)\right| \\
&= \frac{1}{2} \sum_{y \in \Omega'} \left|A(f^{-1}(y)) - B(f^{-1}(y))\right| \\
&= \frac{1}{2} \sum_{y \in \Omega'} \left| \left(\sum_{x \in f^{-1}(y)} A(x) \right) - \left(\sum_{x \in f^{-1}(y)} B(x) \right) \right|
\end{aligned}$$

Proof Continued

- Continuing the manipulation:

$$\begin{aligned}\mathrm{SD}\left(f(A), f(B)\right) &= \frac{1}{2}\sum_{y\in\Omega'}\left|\left(\sum_{x\in f^{-1}(y)} A(x)\right) - \left(\sum_{x\in f^{-1}(y)} B(x)\right)\right| \\ &\leqslant \frac{1}{2}\sum_{y\in\Omega'}\sum_{x\in f^{-1}(y)} |A(x) - B(x)| \\ &= \frac{1}{2}\sum_{x\in\Omega} |A(x) - B(x)| = \mathrm{SD}\left(A, B\right)\end{aligned}$$

- Think: When does equality hold?

Distinguishing Experiment

For two distribution $A^{(0)}$ and $A^{(1)}$ consider the following experiment between an honest challenge $\mathcal{H}$ and an adversary $\mathcal{A}$:

- The honest challenger samples $b \xleftarrow{\$} \{0,1\}$, samples $s \xleftarrow{\$} A^{(b)}$ and sends s to the adversary $\mathcal{A}$
- The adversary $\mathcal{A}$ returns $\widetilde{b}$
- The honest adversary outputs $z = 1$ if and only if $b = \widetilde{b}$

Intuition: The adversary is trying to guess the hidden bit b. If the distributions $A^{(0)}$ and $A^{(1)}$ are dissimilar, then it should be easy for (some) $\mathcal{A}$ to distinguish them. If the distributions $A^{(0)}$ and $A^{(1)}$ are similar, then (any) $\mathcal{A}$ should not be able to distinguish them. Note that (as we had seen earlier) it is easy to achieve $\Pr[z=1] = 1/2$. The advantage of the adversary $\mathcal{A}$ is the probability of $\Pr[z=1]$ beyond $1/2$, i.e. $|\Pr[z=1] - 1/2|$

Examples

- Suppose $A^{(0)}$ and $A^{(1)}$ are identical distributions. Then $\mathrm{SD}\left(A^{(0)}, A^{(1)}\right) = 0$ and the advantage of any adversary is 0
- Suppose $A^{(0)}$ and $A^{(1)}$ are mutually disjoint probabilities, i.e. $\mathrm{SD}\left(A^{(0)}, A^{(1)}\right) = 1$. In this case, there exists an adversary who can ensure $\Pr[z = 1] = 1$, i.e. advantage $1/2$

Relation of Advantage to SD

Claim

The advantage of an adversary is at most $\mathrm{SD}\left(A^{(0)}, A^{(1)}\right)/2$.

- Suppose the adversary sees sample s. Then the best strategy of the adversary is:
  - Output $\widetilde{b} = 0$ if $A^{(0)}(s) > A^{(1)}(s)$
  - Output $\widetilde{b} = 1$ if $A^{(1)}(s) > A^{(0)}(s)$
  - Output any $\widetilde{b}$ if $A^{(0)}(s) = A^{(1)}(s)$

  The probability of $z = 1$ and the sample is s for this algorithm is: $\Pr[b = \widetilde{b}] \cdot \Pr[A^{(\widetilde{b})}(s)] = \max\{A^{(0)}(s), A^{(1)}(S)\}/2$.
- Overall $\Pr[z = 1]$ is

$$\sum_{i \in \Omega} \max\{A^{(0)}(i), A^{(1)}(i)\}/2 = \left(1 + \mathrm{SD}\left(A^{(0)}, A^{(1)}\right)\right)/2$$