Lemma 26.1. *For* $0 \le r \le 1$,

$$\inf_{0\le\lambda\le 1} e^{\frac{1}{4}(1-\lambda)^2} r^{-\lambda} \le 2 - r.$$

Proof. Taking *log*, we need to show

$$\inf_{0\le\lambda\le 1}\left(\frac{1}{4}(1-\lambda)^2 - \lambda\log r - \log(2-r)\right) \le 0.$$

Taking derivative with respect to λ,

$$-\frac{1}{2}(1-\lambda) - \log r = 0$$

$$\lambda = 1 + 2\log r \le 1$$

$$0 \le \lambda = 1 + 2\log r$$

Hence,

$$e^{-1/2} \le r.$$

Take

$$\lambda = \begin{cases} 1 + 2\log r & e^{-1/2} \le r \\ 0 & e^{-1/2} \ge r \end{cases}$$

Case a): $r \le e^{-1/2}$, $\lambda = 0$

$\frac{1}{4} - \log(2-r) \le 0 \;\Leftrightarrow\; r \le 2 - e^{\frac{1}{4}}$. $\quad e^{-1/2} \le 2 - e^{\frac{1}{4}}$.

Case a): $r \ge e^{-1/2}$, $\lambda = 1 + 2\log r$

$$(\log r)^2 - \log r - 2(\log r)^2 - \log(2-r) \le 0$$

Let

$$f(r) = \log(2-r) + \log r + (\log r)^2.$$

Is $f(r) \ge 0$? Enough to prove $f'(r) \le 0$. Is

$$f'(r) = -\frac{1}{2-r} + \frac{1}{r} + 2\log r \cdot \frac{1}{r} \le 0.$$

$$rf'(r) = -\frac{r}{2-r} + 1 + 2\log r \le 0.$$

Enough to show $(rf'(r))' \ge 0$:

$$(rf'(r))' = \frac{2}{r} - \frac{2-r+r}{(2-r)^2} = \frac{2}{r} - \frac{2}{(2-r)^2}.$$

□

Let $\mathcal{X}$ be a set (space of examples) and P a probability measure on $\mathcal{X}$. Let $x_1, \ldots, x_n$ be i.i.d., $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $P^n = P \times \ldots \times P$.

Consider a subset $A \in \mathcal{X}^n$. How can we define a distance from $x \in \mathcal{X}^n$ to A? Example: hamming distance between two points $d(x, y) = \sum I(x_i \neq y_1)$.

We now define *convex hull distance.*

Definition 26.1. *Define $V(A, x)$, $U(A, x)$, and $d(A, x)$ as follows:*

(1) $V(A, x) = \{(s_1, \ldots, s_n) : s_i \in \{0, 1\}, \exists y \in A \text{ s.t. if } s_i = 0 \text{ then } x_i = y_i\}$

$$\begin{array}{rcccc} x = (& x_1, & x_2, & \ldots, & x_n) \\ & = & \neq & \ldots & = \\ y = (& y_1, & y_2, & \ldots, & y_n) \\ s = (& 0, & 1, & \ldots, & 0) \end{array}$$

Note that it can happen that $x_i = y_i$ but $s_i \neq 0$.

(2) $U(A, x) = conv\, V(A, x) = \{\sum \lambda_i u^i,\ u^i = (u_1^i, \ldots, u_n^i) \in V(A, x),\ \lambda_i \geq 0,\ \sum \lambda_i = 1\}$

(3) $d(A, x) = \min_{u \in U(A,x)} |u|^2 = \min_{u \in U(A,x)} \sum u_i^2$

Theorem 26.1.

$$\mathbb{E} e^{\frac{1}{4} d(A,x)} = \int e^{\frac{1}{4} d(A,x)} dP^n(x) \leq \frac{1}{P^n(A)}$$

and

$$P^n(d(A, x) \geq t) \leq \frac{1}{P^n(A)} e^{-t/4}.$$

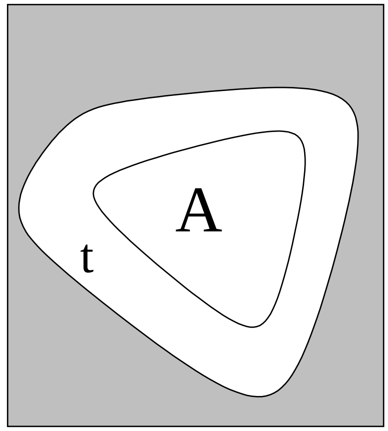
Proof. Proof is by induction on n.

$\mathbf{n = 1}$:

$$d(A, x) = \begin{cases} 0, & x \in A \\ 1, & x \notin A \end{cases}$$

Hence,

$$\int e^{\frac{1}{4} d(A,x)} dP^n(x) = P(A) \cdot 1 + (1 - P(A)) e^{\frac{1}{4}} \leq \frac{1}{P(A)}$$



because

$$e^{\frac{1}{4}} \le \frac{1 + P(A)}{P(A)}.$$

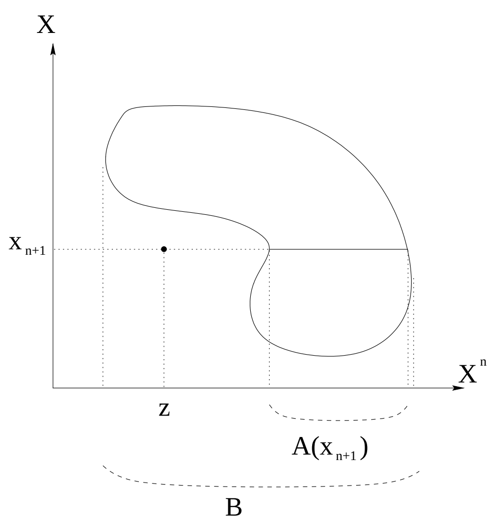
$\mathbf{n \to n+1}$:

Let $x = (x_1, \ldots, x_n, x_{n+1}) = (z, x_{n+1})$. Define

$$A(x_{n+1}) = \{(y_1, \ldots, y_n) : \ (y_1, \ldots, y_n, x_{n+1}) \in A\}$$

and

$$B = \{(y_1, \ldots, y_n) : \ \exists y_{n+1}, \ (y_1, \ldots, y_n, y_{n+1}) \in A\}$$



One can verify that

$$s \in U(A(x_{n+1}, z)) \Rightarrow (s, 0) \in U(A, (z, x_{n+1}))$$

and

$$t \in U(B,z) \Rightarrow (t,1) \in U(A,(z,x_{n+1})).$$

Take $0 \leq \lambda \leq 1$. Then

$$\lambda(s,0) + (1-\lambda)(t,1) \in U(A,(z,x_{n+1}))$$

since $U(A,(z,x_{n+1}))$ is convex. Hence,

$$\begin{aligned} d(A,(z,x_{n+1})) = d(A,x) &\leq |\lambda(s,0)+(1-\lambda)(t,1)|^2 \\ &= \sum_{i=1}^{n}(\lambda s_i + (1-\lambda)t_i)^2 + (1-\lambda)^2 \\ &\leq \lambda \sum s_i^2 + (1-\lambda)\sum t_i^2 + (1-\lambda)^2 \end{aligned}$$

So,

$$d(A,x) \leq \lambda d(A(x_{n+1}),z) + (1-\lambda)d(B,z) + (1-\lambda)^2.$$

Now we can use induction:

$$\int e^{\frac{1}{4}d(A,x)} dP^{n+1}(x) = \int_{\mathcal{X}}\int_{\mathcal{X}^n} e^{\frac{1}{4}d(A,(z,x_{n+1}))} dP^n(z)dP(x_{n+1}).$$

Then inner integral is

$$\begin{aligned} \int_{\mathcal{X}^n} e^{\frac{1}{4}d(A,(z,x_{n+1}))} dP^n(z) &\leq \int_{\mathcal{X}^n} e^{\frac{1}{4}\left(\lambda d(A(x_{n+1}),z)+(1-\lambda)d(B,z)+(1-\lambda)^2\right)} dP^n(z) \\ &= e^{\frac{1}{4}(1-\lambda)^2}\int e^{\left(\frac{1}{4}d(A(x_{n+1}),z)\right)\lambda + \left(\frac{1}{4}d(B,z)\right)(1-\lambda)} dP^n(z) \end{aligned}$$

We now use *Hölder*'s inequality:

$$\int fg dP \leq \left(\int f^p dP\right)^{1/p}\left(\int g^q dP\right)^{1/q} \text{ where } \frac{1}{p}+\frac{1}{q} = 1$$

$$
\begin{aligned}
& e^{\frac{1}{4}(1-\lambda)^2} \int e^{\left(\frac{1}{4}d(A(x_{n+1}),z)\right)\lambda + \left(\frac{1}{4}d(B,z)\right)(1-\lambda)} dP^n(z) \\
& \leq e^{\frac{1}{4}(1-\lambda)^2} \left(\int e^{\frac{1}{4}d(A(x_{n+1}),z)} dP^n(z) \right)^{\lambda} \left(e^{\frac{1}{4}d(B,z)} dP^n(z) \right)^{1-\lambda} \\
& \leq \text{(by ind. hypoth.)} \quad e^{\frac{1}{4}(1-\lambda)^2} \left(\frac{1}{P^n(A(x_{n+1}))} \right)^{\lambda} \left(\frac{1}{P^n(B)} \right)^{1-\lambda} \\
& = \frac{1}{P^n(B)} e^{\frac{1}{4}(1-\lambda)^2} \left(\frac{P^n(A(x_{n+1}))}{P^n(B)} \right)^{-\lambda}
\end{aligned}
$$

Optimizing over $\lambda \in [0, 1]$, we use the Lemma proved in the beginning of the lecture with

$$
0 \leq r = \frac{P^n(A(x_{n+1}))}{P^n(B)} \leq 1.
$$

Thus,

$$
\frac{1}{P^n(B)} e^{\frac{1}{4}(1-\lambda)^2} \left(\frac{P^n(A(x_{n+1}))}{P^n(B)} \right)^{-\lambda} \leq \frac{1}{P^n(B)} \left(2 - \frac{P^n(A(x_{n+1}))}{P^n(B)} \right).
$$

Now, integrate over the last coordinate. When averaging over x_{n+1}, we get measure of A.

$$
\begin{aligned}
\int e^{\frac{1}{4}d(A,x)} dP^{n+1}(x) &= \int_{\mathcal{X}} \int_{\mathcal{X}^n} e^{\frac{1}{4}d(A,(z,x_{n+1}))} dP^n(z) dP(x_{n+1}) \\
&\leq \int_{\mathcal{X}} \frac{1}{P^n(B)} \left(2 - \frac{P^n(A(x_{n+1}))}{P^n(B)} \right) dP(x_{n+1}) \\
&= \frac{1}{P^n(B)} \left(2 - \frac{P^{n+1}(A)}{P^n(B)} \right) \\
&= \frac{1}{P^{n+1}(A)} \frac{P^{n+1}(A)}{P^n(B)} \left(2 - \frac{P^{n+1}(A)}{P^n(B)} \right) \\
&\leq \frac{1}{P^{n+1}(A)}
\end{aligned}
$$

because $x(2-x) \leq 1$ for $0 \leq x \leq 1$.

$\square$