1) **HYPOTHESIS TESTING**

   - **General problem**: Two probability distributions $P_1(x), P_2(x)$ on finite set $X$ are null and are alternative. We observe $n$ iid samples, $x_1, \ldots, x_n$, drawn from an unknown distribution. The goal is generally to determine which distribution sample came from.

2) **TOTAL VARIATION DISTANCE**

   - **Pre requisite to successful hypothesis testing**: There is a way to measure distance between distributions.
   - **Two distributions**: $A_1, A_2, \ldots, A_n$ for $\Omega$ = $\Omega_1$
     
     $$B_1, B_2, \ldots, B_n$$

     $$\text{TVD}(A, B) = \frac{1}{2} \| A - B \|_1 = \frac{1}{2} \sum_{x \in \Omega} |A(x) - B(x)|$$

   - **Intuition**: TVD is best distinguishing probability
     
     The supremum over all subsets of $\Omega$ of $A(x) - B(x)$ [i.e., largest possible difference between the probabilities that two probability distributions can assign to the same event].

   - **More concretely**: Null hypothesis $H_0 = A$
     
     Alternate hypothesis $H_a = B$

     - Have algorithm 1 whose job is, given samples, outputs after each sample if thinks sample came from distribution $A$ or distribution $B$

   - 2 common types of error: Type I = False $\Theta$ = Incorrectly reject true $H_0$
     
     Type II = False $\bar{\Theta}$ = Fail to reject false $H_0$

   - Let's say I reject $A$ when event $\bar{Z}$ occurs:
     
     Type I error + Type II error $= A(\bar{Z}) + B(\bar{Z})$

     $$= A(\bar{Z}) + (1 - B(\bar{Z}))$$

     $$= 1 + (A(\bar{Z}) - B(\bar{Z}))$$

     $$\geq 1 + \inf_{A} [A(\bar{Z}) - B(\bar{Z})]$$

     $$= 1 - \sup_{A} [A(\bar{Z}) - B(\bar{Z})]$$

     $$= 1 - \text{TVD}(A, B)$$

   - Missing step? $\text{TVD}(A, B) = \sup_A |A - B|$

   - $\sup(1 - B(\bar{Z})) = \sup_A (A - B)$

   - Why not just use KL?
     
     - Non symmetric = KL
     
     - $1 - \text{TVD}$ is lower bound of Type I and Type II error rates
     
     - Many other distances, but this one is most distinguishing
3. **TVD Behavior**

- If $A \neq B$, then $\lim_{n \to \infty} \text{TVD}(A^n, B^n) \to 1$

- but $\text{TVD}(A^{k+1}, B^{k+1})$ can be equal to $\text{TVD}(A^k, B^k)$

4. **Pinkevich's Inequality**

Recall KL divergence $D(P \parallel Q) = \sum x p(x) \log_2 \frac{p(x)}{q(x)}$

if $P$ and $Q$ are discrete distributions, then

| $D(P \parallel Q) \geq \frac{1}{2 \ln(2)} \| P - Q \|^2_1$ |

so we can use KL to upper bound the TVD: $\text{TVD}(P, Q) \leq \frac{1}{2 \sqrt{\ln(2)}} D(P \parallel Q)$

**Proof:** I'll prove a special case of coin flip $\text{Bernoulli} \times 2$

Let $P = \{0 \ p, 1 \ 1-p \}$ and $Q = \{0 \ wp, 1-q \}$

Assume $p \geq q$ (but similar steps allow us to prove the other case)

$D(P \parallel Q) = p \log \frac{p}{q_0} + (1-p) \log \frac{1-p}{q_0} + \frac{1}{2 \ln(2)} (2p-q)^2$

where $f(p, q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{q} - \frac{1}{2 \ln(2)} (2p-q)^2$

Let $\frac{\partial f}{\partial p} = -\frac{q}{\ln(2)} \left( \frac{1}{2} \ln(2) - 1 \right) \leq 0$ if $q = \frac{1}{2}$ otherwise negative, since $p-q \leq 0, \theta \text{ sign, ( ) is negative}$

So $f(p, q) \geq 0$ if $p \geq q$ so $D(P \parallel Q) \geq \frac{1}{2 \ln(2)} \| P - Q \|^2_1$ for this special case.

**Chain Rule of KL**

$KL(P \parallel Q) = KL(P \parallel R) + KL(R \parallel Q)$

**Sketch:** want to prove for any distributions $P$ and $Q$, calculate $\| P - Q \|_1$, and simplify to $\| P - Q \|_1$, where $P = \text{Bernoulli}(\frac{1}{2} P(x))$ (these two will be equal)

then use chain rule of KL divergence and the fact that KL is non-negative.
APPLICATION of PINSKER: lower bound of coin samples needed to distinguish two coins with slightly different biases

Let $H$ = heads, $T$ = tails
we are given one of two coins:
\[
P = \{ \begin{array}{ll} 0 & \text{wp } \frac{1}{2} + \epsilon \\ 1 & \text{wp } \frac{1}{2} - \epsilon \end{array}, \quad Q = \{ \begin{array}{ll} 0 & \text{wp } \frac{1}{2} + \epsilon \\ 1 & \text{wp } \frac{1}{2} - \epsilon \end{array} \}
\]
we have our algorithm $A(x_1, \ldots, x_m) \rightarrow \{0, 1\}$
takes output of $m$ coin tosses and decides which coin the
 tosses came from

assume our algorithm is good, i.e.
\[
\Pr(A(x) = 0) \geq \frac{9}{10} \quad \forall x \in \{\text{p}(x) = 1\}
\]
goal: derive a lower bound for $m$ (without knowing $A$)

rewrite assumption conditions in terms of expectations:
\[
\E_{x \sim \text{p}(x)}[A(x)] = \frac{1}{10} \quad \E_{x \sim \text{q}(x)}[A(x)] = \frac{9}{10}
\]
so $\E_{x \sim \text{p}(x)}[A(x)] - \E_{x \sim \text{q}(x)}[A(x)] \geq \frac{8}{10}$

apply this lemma: $\tilde{p}$ and $\tilde{q}$ are distributions on some universal pedantic world $U$

let $f: U \rightarrow \{0, 1\}$ and $f$ represents a discrete upper bound

**Proof:** rewrite using definition of expected value/lotus
\[
\left| \E_{x \sim \tilde{p}(x)} f(x) - \E_{x \sim \tilde{q}(x)} f(x) \right| = \left| \sum_{x} \tilde{p}(x) f(x) - \sum_{x} \tilde{q}(x) f(x) \right|
\]
\[
= \left| \sum_{x} \left( \tilde{p}(x) - \tilde{q}(x) \right) f(x) - \frac{\E}{2} (\sum_{x} \tilde{p}(x) - \tilde{q}(x)) \right|
\]
\[
\leq \sum_{x} \left| \tilde{p}(x) - \tilde{q}(x) \right| |f(x) - \frac{\E}{2}|
\]
\[
\leq \frac{\E}{2} \left\| \tilde{p} - \tilde{q} \right\|_1
\]
now let $f$ be A, $\tilde{f} = p^m$, $\tilde{A} = q^m$

$$\| p^m - q^m \|_1 \geq 2 \left| \frac{1}{x \in p^m} A(x) - \frac{1}{x \in q^m} A(x) \right| = \frac{\delta}{5}$$

recall from lecture 3 that:

$$m D(\Pi(\Omega) || p^m || q^m) \geq \frac{1}{2m^2} \left( \frac{\delta}{5} \right)^2$$

by Riesz's inequality

$$m \geq \frac{1}{2} \frac{1}{\ln 2} \frac{1}{\Pi(\Pi(\Omega))} \left( \frac{\delta}{5} \right)^2$$

now last thing is to bound $D(\Pi(\Omega))$:

$$D(\Pi(\Omega)) = \frac{1}{2} \log \left( \frac{\frac{1-E}{E}}{\frac{1+E}{E}} \right) = \frac{1}{2} \log \left( \frac{1+2E}{1-2E} \right)$$

$$\leq \frac{\frac{E}{1-2E}}{\ln 2} \ln \left( 1 + \frac{4E}{1-2E} \right)$$

$$\leq \frac{4E^2}{(1-2E) \ln 2}$$

since $\ln (1+x) \leq ex$

if we assume $E \approx \frac{1}{4}$ i.e. is small

$$D(\Pi(\Omega)) \leq \frac{8E^2}{(1-2E) \ln 2}$$

so

$$m \geq \frac{1}{2} \frac{1}{\ln 2} \frac{1}{\Pi(\Pi(\Omega))} \left( \frac{\delta}{5} \right)^2 \geq \frac{4E^2}{254}$$

can show this is upto constants using Chernoff bound
Lower Bound on TVD

"Direct Product" Lemma: If $X$ and $Y$ are distributions such that $\text{TVD}(X, Y) = \delta$
then for $k \in \mathbb{N}$,
$$1 - 2e^{-k\delta^2/2} \leq \text{TVD}(X_k, Y_k)$$

Proof: Recall that TVD is the best distinguishing probability so there exists some set $S$ such that
$$\mathbb{P}(X \in S) - \mathbb{P}(Y \in S) = \delta$$

Let $p = \mathbb{P}(Y \in S)$ so $\mathbb{P}(X \in S) = \delta + p$

Then in $k$ samples of $X$, expected # of those that lie in $S$ are $(\delta + p)k$ and similarly for $Y$ is $pk$

We now apply Chernoff bounds:
$$\mathbb{P}(\text{at least} \ (p + \frac{\delta}{2})k \text{ components of } Y_k \text{ lie in } S) \geq \exp(-k\delta^2/2)$$
$$\mathbb{P}(\text{at most} \ (p + \frac{\delta}{2})k \text{ components of } X_k \text{ lie in } S) \geq \exp(-k\delta^2/2)$$

Let $S'$ be the set of $k$-tuples that contain more than $(p + \frac{\delta}{2})k$ components that lie in $S$, then:
$$\text{TVD}(X_k, Y_k) \geq \mathbb{P}(X_k \in S') - \mathbb{P}(Y_k \in S')$$
$$\geq 1 - 2e^{-k\delta^2/2}$$

So we can lower bound TVD and upper bound TVD for Bernoulli coin flip.

Direct product lemma

Chernoff inequality