1 Review and formulas

- Le Cam’s identity: \( P_{c}^* = \frac{1}{2} - \frac{1}{2}||P_1 - P_2||_{TV} \)
- Pinsker’s inequality: \( \frac{2}{\ln 2}||P_1 - P_2||_{TV}^2 \leq D(P_1 \| P_2) \)

2 Proving Pinsker’s inequality

Take two Bernoulli distributions \( P_1, P_2 \), where \( P_1(X = 1) = p, P_2(X = 1) = q \). With some elementary substitution, we can manipulate the right side of the equation:

\[
D(P_1 \| P_2) = D(p \| q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q}
\]

We can also change the left side:

\[
\frac{2}{\ln 2}||P_1 - P_2||_{TV}^2 = \frac{1}{2 \ln 2}||P_1 - P_2||_{\ell_1}^2 = \frac{1}{2 \ln 2} (2|p - q|)^2 = \frac{2(p - q)^2}{\ln 2}
\]

Our new goal is to show that the new right side minus the left side is \( \geq 0 \), or: \( f(p, q) = p \ln \frac{p}{q} + (1 - p) \ln \frac{1 - p}{1 - q} - 2(p - q)^2 \geq 0 \), as this is equivalent to proving Pinsker’s inequality.

If we differentiate \( f \) with respect to \( q \), we get:

\[
\frac{df}{dq} = -\frac{p}{q} + (1 - p) \frac{1}{1 - q} + 4(p - q)
\]

\[
= -\frac{p(1 - q) + q(1 - p)}{q(1 - q)} + 4(p - q)
\]

\[
= -\frac{p + q}{q(1 - q)} + 4(p - q)
\]

\[
= (p - q)(4 - \frac{1}{q(1 - q)})
\]

Is this differentiation increasing or decreasing? Since \( 0 \leq q \leq 1 \), the maximum value \( q(1 - q) \) can take is \( \frac{1}{4} \), which occurs when \( q = \frac{1}{2} \). So, the minimum value of \( \frac{1}{q(1 - q)} \) is 4, so \( 4 - \frac{1}{q(1 - q)} \) will always be negative. Thus, whether this differentiation is increasing or decreasing depends on \( p - q \).

If \( p \geq q \) then \( \frac{df}{dq} \leq 0 \). If \( p \leq q \) then \( \frac{df}{dq} \geq 0 \). Thus, \( f(p, q) \) looks like an upside down parabola which is lowest when \( p = q \) and is always \( \geq 0 \). Therefore, since \( f(p, q) \geq 0 \), \( \frac{2}{\ln 2}||P_1 - P_2||_{TV}^2 \leq D(P_1 \| P_2) \), which proves Pinsker’s inequality for Bernoulli random variables.
2.1 Implications
We can use Pinsker’s inequality to show this:

\[ D(p + \epsilon||p) \geq \frac{1}{2 \ln 2}(2\epsilon)^2 = \frac{2\epsilon^2}{\ln 2} \]

This is the Chernoff bound.

3 Neyman-Pearson test

3.1 Binary hypothesis testing
Any binary hypothesis testing divides sample space into 2 parts, creating estimator \( g(X) \) PUT PIC HERE (Instructor’s comment: Poor work by scribe)

What is the probability of error for \( g(X) \)?

\( H_1 \) chosen, if \( X \in A \); error = \( P_1(A^c) \)
\( H_2 \) chosen, if \( X \in A^c \); error = \( P_2(A) \)

Fix one \( P(A^c) \) and then minimize the other

3.2 Neyman-Pearson test
Define: \( A(T) = \{ x \in \mathcal{X} : \frac{P_1(x)}{P_2(x)} \geq T \} \), where \( A(T) \) is the decision region, and \( T \) is threshold

3.3 Proving Neyman-Pearson’s claim
Proof: Neyman-Pearson optimality

\( P_1(A^c) \overset{\Delta}{=} \alpha \), where \( \alpha \) is probability of error
\( P_2(A) \overset{\Delta}{=} \beta \), where \( \beta \) is probability of error

In the Neyman-Pearson test, as \( \rightarrow A(T) \),
\( P_1(A(T)) \overset{\Delta}{=} \alpha^* \)
\( P_2(A(T)) \overset{\Delta}{=} \beta^* \)

Suppose there is another test with decision region \( B, B^c \)
The claim of the Neyman-Pearson lemma is that if
\( \alpha < \alpha^* \), then \( \beta > \beta^* \).

If you can design any test that has better probability of error for the first term, the test will have a worse probability of error for the second term.

another picture here (Instructor’s comment: Poor work)
\( \forall x \in X \), we have

\[
\mathbb{1}_A(x) = \begin{cases} 
1, & \text{if } x \in A \\
0, & \text{otherwise}
\end{cases}
\]

\[
\mathbb{1}_B(x) = \begin{cases} 
1, & \text{if } x \in B \\
0, & \text{otherwise}
\end{cases}
\]
Show that \((\mathbb{I}_A(x) - \mathbb{I}_B(x))(P_1(x) - TP_2(x)) \geq 0 \forall x \in \mathcal{X}\)

**Case 1:** \(x \in A\),

\[
(\mathbb{I}_A(x) - \mathbb{I}_B(x))(P_1(x) - TP_2(x)) = (1 - \mathbb{I}_B(x))(\text{positive value}) \geq 0
\]

\(\mathbb{I}_B(x)\) is 1 or 0.

**Case 2:** \(x \in A^c\)

\[
(0 - \mathbb{I}_B(x))(\text{negative value}) \geq 0
\]

\(\mathbb{I}_B(x)\) is a negative value or 0.

Therefore, \((\mathbb{I}_A(x) - \mathbb{I}_B(x))(P_1(x) - TP_2(x)) \geq 0 \forall x \in \mathcal{X}\)

\[
\sum_{x \in X}(\mathbb{I}_A(x)P_1(x) - T\mathbb{I}_A(x)P_2(x) - \mathbb{I}_B(x)P_1(x) + T\mathbb{I}_B(x)P_2(x)) \geq 0
\]

\[
\sum_{x \in A} P_1(x) + \sum_{x \in B} P_1(x) - T \sum_{x \in A} P_2(x) + T \sum_{x \in B} P_2(x) \geq 0
\]

\[P_1(A) - P_1(B) - TP_2(A) + TP_2(B) \geq 0\]

\[P_1(A) - P_1(B) \geq T(P_2(A) - P_2(B))\]

\[(1 - \alpha^*) - (1 - \alpha) \geq T(\beta^* - \beta)\]

\[\alpha - \alpha^* \geq T(\beta^* - \beta)\]

Therefore, if \(\alpha < \alpha^* \Rightarrow \beta > \beta^*\).

# 4 Comparing Neyman-Pearson and Bayes’ tests

## 4.1 Bayes’ test

The Bayes test uses Bayes rule to derive an estimator based on the prior likelihood of a hypothesis. Say we have two probability distributions \(P_1, P_2,\) and \(H_1\) is the hypothesis that a sample came from \(P_1\).

\[
\max_{i \in \{1, 2\}} P(H_i | X = x) = \max_{i \in \{1, 2\}} \frac{P(H_i, X = x)}{P(X = x)}
\]

\[
= \max_{i \in \{1, 2\}} P(X = x | H_i) \frac{P(H_i)}{P(X = x)}
\]

\[
= \max_{i \in \{1, 2\}} P(X = x | H_i) P(H_i)
\]

If \(P(x|H_1)P(H_1) \geq P(x|H_2)P(H_2)\), then our estimator \(g(x) = 1\), which means it chooses \(P_1\). Equivalently, if \(P_1(x)P(H_1) > P_2(x)P(H_2)\). We can describe the Bayes decision region like so:

\[
A = \{x | \frac{P_1(x)}{P_2(x)} \geq \frac{P(H_2)}{P(H_1)}\}
\]

This shows us that if we want to use the Bayes test, we need to know the prior probabilities, which may not be available. Neyman-Pearson test is more flexible because it doesn’t require priors and because you can choose a threshold \(T\) for Neyman-Pearson. Thus, more people tend to prefer it.
5 Neyman-Pearson with multiple samples

Say we have multiple samples, \(x_1, \ldots, x_n\). We can reformat the Neyman-Pearson test like this:

\[
A(T) = \{(x_1, \ldots, x_n) \in \mathcal{X}^n | \frac{P_1(x_1, \ldots, x_n)}{P_2(x_1, \ldots, x_n)} > T\}
\]

If all samples are independent, then:

\[
= \{(x_1, \ldots, x_n) \in \mathcal{X}^n | \prod_{i=1}^{n} \frac{P_1(x_i)}{P_2(x_i)} \geq T\}
\]

If we take the log, the product becomes a sum, and the test becomes equivalent to the log likelihood ratio:

\[
A(T) = \{(x_1, \ldots, x_n) \in \mathcal{X}^n | \sum_{i=1}^{n} \log \frac{P_1(x_i)}{P_2(x_i)} \geq \log T\}
\]

If we define a function \(\lambda(x) = \log \frac{P_1(x)}{P_2(x)}\), we can rewrite the log likelihood ratio:

\[
\sum_{i=1}^{n} \lambda(x_i) \geq \log T
\]

\[
\frac{1}{n} \sum_{i=1}^{n} \lambda(x_i) \geq \frac{\log T}{n}
\]