The Complexity of Distinguishing Distributions

Serge Vaudenay

http://lasecwww.epfl.ch/
1 From Statistical Distance to Chernoff Information

2 Applications

3 Further Extensions
1. From Statistical Distance to Chernoff Information
   - A Common Cryptographic Problem
   - Hypothesis Testing
   - Best Advantage with Single Sample
   - Chernoff Information
   - Chernoff Bound
   - Approximations of the Chernoff Information
   - Consequence of the Sanov Theorem for Same Support
   - Application to Composite Hypothesis Testing

2. Applications

3. Further Extensions
Indistinguishability

Problem: say if samples follow distribution $P_0$ or $P_1$
**Advantage**

**Definition**

Two samplable distributions $P_0$ and $P_1$ are $(q, \varepsilon)$-indistinguishable if for any algorithm $\mathcal{A}$ taking $q$ iid random variables $x_1, \ldots, x_q$ following $P$ we have

$$|\text{Adv}_{\mathcal{A}}(P_0, P_1)| \leq \varepsilon$$

where

$$\text{Adv}_{\mathcal{A}}(P_0, P_1) = \Pr[\mathcal{A} \rightarrow 1|P = P_1] - \Pr[\mathcal{A} \rightarrow 1|P = P_0]$$

A notion of distance between $P_0$ and $P_1$:  

$$\text{distance}_q(P_0, P_1) = \min_{\text{distinguisher limited to } q} |\Pr[\mathcal{A} \rightarrow 1|P = P_1] - \Pr[\mathcal{A} \rightarrow 1|P = P_0]|$$
Applications

- pseudorandom number generator
  break a PRNG $\Rightarrow$ distinguish from an ideal RNG

- block cipher and stream cipher cryptanalysis
  distinguish biased bits in known plaintext-ciphertexts

- semantic security of public-key cryptography
  distinguish between the encryption of two known plaintexts

- commitment, zero-knowledge, etc
Example: Decisional Diffie-Hellman Problem

Assume a group \((\mathbb{Z}_p^*, \text{an elliptic curve, ...})\) generated by some \(g\)

Alice

- pick \(x\) at random, \(X \leftarrow g^x\)
- \(K \leftarrow Y^x\)

Bob

- \(X \xrightarrow{\ } X\)
- \(Y \xleftarrow{\ } Y\)
- pick \(y\) at random, \(Y \leftarrow g^y\)
- \(K \leftarrow X^y\)

\((K = g^{xy})\)

Problem: given a single sample \((X, Y)\) and a candidate \(\kappa\) for \(K\) tell if \((X, Y, \kappa)\) follows distribution of \((X, Y, K)\) or the one of \((X, Y, Z)\) with \(Z\) random
Example: Block Cipher Cryptanalysis

Problem: given many samples \((X, Y)\) and a candidate value \(\kappa\) for \(K_{\text{last}}\), tell if \((X, \text{round}_{\kappa}^{-1}(Y))\) follows distribution of \((X, \text{core}_K(X))\) or some garbage distribution.
Example: Semantic Security

**Definition**

Cryptosystem is IND-CPA secure if \( \Pr[\text{win}] - \frac{1}{2} \) is negligible for any such adversary.

Problem of the adversary: tell if \( y \) (single sample) follows distribution of \( \text{Enc}(x_0) \) or the one of \( \text{Enc}(x_1) \)
1 From Statistical Distance to Chernoff Information
   • A Common Cryptographic Problem
   • Hypothesis Testing
     • Best Advantage with Single Sample
     • Chernoff Information
     • Chernoff Bound
     • Approximations of the Chernoff Information
     • Consequence of the Sanov Theorem for Same Support
     • Application to Composite Hypothesis Testing

2 Applications

3 Further Extensions
Hypothesis Testing

Problem

Given a source producing random variables, decide upon several hypotheses.

Example:

- iid random variables following either

  **Hypothesis** $H_0$: variables follow distribution $P_0$
  **Hypothesis** $H_1$: variables follow distribution $P_1$

- iid random variables following either

  **Hypothesis** $H_0$: variables follow distribution $P_0$
  **Hypothesis** $H_1$: variables follow distribution in $\{P_1, \ldots, P_n\}$
Two Approaches

- **Frequentist approach**
  Consider two types of errors
  
  **type I error:** $\alpha = \Pr[\mathcal{A} \rightarrow 1 | P_0]$
  
  **type II error:** $\beta = \Pr[\mathcal{A} \rightarrow 0 | P_1]$

- **Bayesian approach**
  Assign cost to error type (or prior probability to hypotheses)

  $$P_e = \Pr[\mathcal{A} \rightarrow 1 | P_0] \pi_0 + \Pr[\mathcal{A} \rightarrow 0 | P_1] \pi_1$$

  Typical case for crypto: $\pi_0 = \pi_1 = \frac{1}{2}$

  $$\text{Adv}_{\mathcal{A}} = (1 - \beta) - \alpha = 1 - 2P_e$$

  $$1 - \text{Adv}_{\mathcal{A}} = \alpha + \beta = 2P_e$$
Problems for this Lecture

- What is the best way to distinguish two distributions?
- What is the difference between using a single sample or many samples?
- How many samples do we need to distinguish two distributions with significant advantage?
From Statistical Distance to Chernoff Information
- A Common Cryptographic Problem
- Hypothesis Testing
- Best Advantage with Single Sample
- Chernoff Information
- Chernoff Bound
- Approximations of the Chernoff Information
- Consequence of the Sanov Theorem for Same Support
- Application to Composite Hypothesis Testing

Applications

Further Extensions
Best Advantage

Case $q = 1$

- let $\mathcal{A}$ be an arbitrary distinguisher
- w.l.o.g. we can assume it is deterministic (we assume no computational bound)
  → let $\mathcal{A}^{-1}(1)$ be the set of values $x$ such that $\mathcal{A} \rightarrow 1$ when $X = x$
- we have
  $$
  \text{Adv}_{\mathcal{A}} = \sum_{x \in \mathcal{A}^{-1}(1)} (P_1(x) - P_0(x))
  $$
- clearly
  $$
  \text{Adv}_{\mathcal{A}} \leq \sum_{x; P_0(x) \leq P_1(x)} (P_1(x) - P_0(x))
  $$
- we have
  $$
  \sum_{x; P_0(x) \leq P_1(x)} (P_1(x) - P_0(x)) = \frac{1}{2} \sum_x |P_1(x) - P_0(x)|
  $$
Definition (\(= L_1 \) distance)

Given two real functions \( f_0 \) and \( f_1 \) over a discrete set \( \mathcal{Z} \) we define the **statistical distance** \( d(f_0, f_1) \) by

\[
d(f_0, f_1) = \frac{1}{2} \sum_{x \in \mathcal{Z}} |f_1(x) - f_0(x)|
\]

Theorem

*Given two distributions \( P_0 \) and \( P_1 \), all distinguishers using a single sample verify*

\[
\text{Adv}_\mathcal{A} \leq d(P_0, P_1)
\]
Best Distinguisher

input: $x$

threshold: $\tau = 1$

1: $R = \frac{P_0(x)}{P_1(x)}$
2: if $R \leq \tau$ then
3: $b \leftarrow 1$
4: else
5: $b \leftarrow 0$
6: end if

output: $b$

- $R$ is the likelihood ratio

$\text{Adv}_A = d(P_0, P_1)$

- caveat: $\frac{p}{0} = +\infty$
- remark: $\frac{0}{0}$ never occurs
The best possible advantage is obtained by the likelihood ratio test:

\[
\text{output } 1 \iff \frac{\Pr_D[z_1, \ldots, z_n]}{\Pr_{D^*}[z_1, \ldots, z_n]} > 1
\]
Theorem (Neyman-Pearson 1933)

Given two distributions $P_0$ and $P_1$, let $\mathcal{A}$ be a distinguisher with error probabilities $\alpha$ and $\beta$.
Let $\tau$ be a threshold defining a likelihood ratio distinguisher with error probabilities $\alpha^*$ and $\beta^*$.
If $\alpha < \alpha^*$ then $\beta > \beta^*$. we cannot beat both $\alpha^*$ and $\beta^*$ for a distinguisher based on the likelihood ratio
General Case

trick: consider $X = (X_1, \ldots, X_q)$ as a random variable with distribution either $P_0^\otimes q$ or $P_1^\otimes q$

input: $x_1, \ldots, x_q$

threshold: $\tau = 1$

1: $R = \frac{P_0(x_1) \times \cdots \times P_0(x_q)}{P_1(x_1) \times \cdots \times P_1(x_q)}$

2: if $R \leq \tau$ then
3: $b \leftarrow 1$

4: else
5: $b \leftarrow 0$

6: end if

output: $b$
Example: Biased Coin

\[ P_0 = \text{uniform} \quad P_1 = \begin{pmatrix} 1 & 2 \\ \frac{1}{2}(1 + \varepsilon) & \frac{1}{2}(1 - \varepsilon) \end{pmatrix} \]

<table>
<thead>
<tr>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(R)</th>
<th>outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>(\frac{1}{(1 + \varepsilon)^2})</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>(\frac{1}{(1 - \varepsilon)^2})</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(\frac{1}{(1 + \varepsilon)(1 - \varepsilon)})</td>
<td>0</td>
</tr>
</tbody>
</table>

output 1 \(\Longleftrightarrow n_2 \leq n_1\)
Example: Biased Dice

\[ P_0 = \text{uniform} \quad P_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \frac{1}{6} + \varepsilon & \frac{1}{6} + \varepsilon & \frac{1}{6} - \varepsilon & \frac{1}{6} - \varepsilon & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \]

<table>
<thead>
<tr>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( R )</th>
<th>outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>( \frac{1}{6} \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} )</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>( \frac{1}{6} + \varepsilon \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} )</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>( \frac{1}{6} - \varepsilon \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} ) ( \frac{1}{6} \times \frac{1}{6} )</td>
<td>0</td>
</tr>
</tbody>
</table>

output 1 \( \sim \) \( n_4 + n_5 \leq n_1 + n_3 \)
Example: Uniform over Different Supports

\[ P_0 = \text{uniform} \]

\[ P_1 = \begin{pmatrix}
1 & 2 & 3 & 4 & 5 \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0
\end{pmatrix} \]

<table>
<thead>
<tr>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(R)</th>
<th>outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>(\frac{1}{5} \times \frac{1}{5})</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>(\frac{1}{4} \times \frac{1}{4})</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>(\frac{1}{4} \times \frac{1}{4})</td>
<td>0</td>
</tr>
</tbody>
</table>

Output 1 \(\iff\) \(n_5 = 0\)
input: $x_1, \ldots, x_q$
threshold: $\tau = 1$

1: $L = \sum_{i=1}^{q} \log \frac{P_0(x_i)}{P_1(x_i)}$
2: if $L \leq \log \tau$ then
3: $b \leftarrow 1$
4: else
5: $b \leftarrow 0$
6: end if

output: $b$

- $\log 0 = -\infty$
- $\log(\infty) = +\infty$
- we never have $+\infty - \infty$

Adv$_A = \frac{1}{2} \sum_{y \in \mathbb{Z}^q} \left| \Pr_{P_1^\otimes q}[y] - \Pr_{P_0^\otimes q}[y] \right|$
Problem

\[
\text{Adv}_{\mathcal{A}} = \frac{1}{2} \sum_{x_1, \ldots, x_q \in \mathbb{Z}} \left| \Pr_{P_1^{\otimes q}} [x_1, \ldots, x_q] - \Pr_{P_0^{\otimes q}} [x_1, \ldots, x_q] \right|
\]

not very informative about the dependence in terms of \( q \)
for $q \ll 1/d(P_0, P_1)$ the advantage must be negligible:

**Theorem**

For any $q$:

$$d(P_0^\otimes q, P_1^\otimes q) \leq q \times d(P_0, P_1)$$

**Proof.**

$$aa' - bb' = (a - b) \frac{a' + b'}{2} + (a' - b') \frac{a + b}{2}$$

so $|aa' - bb'| \leq |a - b| + |a' - b'|$ thus

$$\frac{1}{2} \sum_{x_1, x_2} |P_1(x_1)Q_1(x_2) - P_0(x_1)Q_0(x_2)| \leq d(P_0, P_1) + d(Q_0, Q_1)$$

and we get $d(P_0 \otimes Q_0, P_1 \otimes Q_1) \leq d(P_0, P_1) + d(Q_0, Q_1)$

apply with $Q_b = P_b^{\otimes (q-1)}$ and iterate
1 From Statistical Distance to Chernoff Information
   - A Common Cryptographic Problem
   - Hypothesis Testing
   - Best Advantage with Single Sample
   - Chernoff Information
   - Chernoff Bound
   - Approximations of the Chernoff Information
   - Consequence of the Sanov Theorem for Same Support
   - Application to Composite Hypothesis Testing

2 Applications

3 Further Extensions
Definitions — i

- **Type** of a sample vector $y = (x_1, \ldots, x_q)$: distribution $P_y$ such that
  \[ P_y(z) = \frac{1}{q} \# \{ i; x_i = z \} \]
  = observed distribution

- **Kullback-Leibler divergence**:

  \[
  D(P_0 || P_1) = \sum_{x \in \text{Supp}(P_0)} P_0(x) \log \frac{P_0(x)}{P_1(x)}
  \]

  always non-negative, 0 iff $P_0 = P_1$
  infinite iff $\text{Supp}(P_0) \not\subseteq \text{Supp}(P_1)$
WARNING

log is in basis 2!
Best Distinguisher in Terms of Type

We have

\[ \log R = \log \frac{P_0(x_1) \times \cdots \times P_0(x_q)}{P_1(x_1) \times \cdots \times P_1(x_q)} = q \sum_{z \in \mathbb{Z}} P_x(z) \log \frac{P_0(z)}{P_1(z)} \]

Let

\[ \Pi = \left\{ P; \sum_{z \in \mathbb{Z}} P(z) \log \frac{P_0(z)}{P_1(z)} \leq 0 \right\} = \{ P; D(P \parallel P_1) \leq D(P \parallel P_0) \} \]

We have

\[ R \leq 1 \iff P_x \in \Pi \]
Definitions — ii

- **Chernoff information:**
  \[
  C(P_0, P_1) = -\log \inf_{0<\lambda<1} f(\lambda)
  \]
  \[
  f(\lambda) = \sum_{x \in \text{Supp}(P_0) \cap \text{Supp}(P_1)} P_0(x)^{1-\lambda} P_1(x)^{\lambda}
  \]

- **Asymptotic equivalence:** \( f(q) \asymp g(q) \) means \( f(q) = g(q) e^{o(q)} \) when \( q \to +\infty \)
Sanov Theorem

**Theorem**

Let $\mathcal{Z}$ be a finite set. Let $P$ be a distribution over $\mathcal{Z}$. Let $\Pi$ be a set of distributions over $\mathcal{Z}$ such that $\Pi = \Pi$. Let $Y$ be a random vector of $q$ iid samples following $P$. We have

$$\Pr[P_Y \in \Pi] = 2^{-qD(\Pi||P)}$$

where $D(\Pi||P) = \inf_{Q \in \Pi} D(Q||P)$.

Interpretation of the $\Pi = \Pi$ hypothesis: $\Pi$ has no isolated point.
Let $\mathcal{Z}$ be a finite set of cardinality $n$.

- distribution $\equiv$ real vector of $n$ coordinates
- consider any norm definition over $\mathbb{R}^n$
  (they define the same topology)
- open sets: union of open balls
  notation: $\Pi$ is the union of open sets included in $\Pi$
- closed sets: complement of open sets
  notation: $\overline{\Pi}$ is the intersection of closed sets containing $\Pi$
- remark: the set of distributions is bounded and topologically closed in a finite vector space $\implies$ it is a compact set

**consequence:** for any closed sets of distribution and any continuous function over this set, $\inf f$ and $\sup f$ are reached in the set
Lemma

Let $\mathcal{Z}$ be a finite set. Let $P_0$ and $P_1$ be two distributions with support of union $\mathcal{Z}$. Let

$$\Pi = \{P; D(P \| P_1) \leq D(P \| P_0)\}$$

Let $Y$ be a random vector of $q$ iid samples following $P_0$. We have

$$\Pr[P_Y \in \Pi] = \left(\inf_{\lambda > 0} \sum_{x \in \text{Supp}(P_0) \cap \text{Supp}(P_1)} P_0(x)^{1-\lambda} P_1(x)^{\lambda}\right)^q$$

$$\alpha = \left(\inf_{\lambda > 0} f(\lambda)\right)^q \quad \beta = \left(\inf_{\lambda < 1} f(\lambda)\right)^q$$
Best Advantage

**Theorem**

Let $\mathcal{Z}$ be a finite set. Let $P_0$ and $P_1$ be two distributions with support of union $\mathcal{Z}$. Let $\text{BestAdv}_q(P_0, P_1)$ be the best advantage for distinguishing $P_0$ from $P_1$ using $q$ samples. We have

$$1 - \text{BestAdv}_q(P_0, P_1) = 2^{-qC(P_0, P_1)}$$

**Proof.** Using the previous result $1 - \text{BestAdv}_q(P_0, P_1)$ expresses as

$$\left( \inf_{\lambda > 0} f(\lambda) \right)^q + \left( \inf_{\lambda < 1} f(\lambda) \right)^q$$

since $f'$ vanishes at most once, $\max(\inf_{-\infty, 1}, \inf_{0, +\infty}) = \inf_{0, 1}$

we need a number of samples $\sim 1/C(P_0, P_1)$
Example: Biased Coin

\[ P_0 = \text{uniform} \quad P_1 = \left( \frac{1}{2}(1 + \varepsilon), \frac{1}{2}(1 - \varepsilon) \right) \]

\[ f(\lambda) = \left( \frac{1}{2} \right)^{1-\lambda} \left( \frac{1}{2} + \frac{\varepsilon}{2} \right)^{\lambda} \]

\[ = \frac{1}{2} \times \left( (1 - \varepsilon)^{\lambda} + (1 + \varepsilon)^{\lambda} \right) \]

minimum reached for \( \lambda \approx \frac{1}{2} + \frac{\varepsilon}{2} \)

\[ C(P_0, P_1) \approx -\log \left( 1 - \frac{\varepsilon^2}{8} \right) \approx \frac{\varepsilon^2}{8 \ln 2} \]

we deduce

\[ \alpha \bullet \beta \bullet 1 - \text{BestAdv}_q \sim e^{-\frac{q}{8} \varepsilon^2} \]

For information: the easy bound was \( \text{BestAdv}_q \leq q \times \frac{\varepsilon}{2} \)
Example: Biased Dice

\[ P_0 = \text{uniform} \quad P_1 = \left( \frac{1}{6} + \varepsilon \quad \frac{1}{6} \quad \frac{1}{6} + \varepsilon \quad \frac{1}{6} - \varepsilon \quad \frac{1}{6} - \varepsilon \quad \frac{1}{6} \right) \]

\[
f(\lambda) = 2 \left( \frac{1}{6} \right)^{1 - \lambda} \left( \frac{1}{6} + \varepsilon \right)^{\lambda} + 2 \left( \frac{1}{6} \right)^{1 - \lambda} \left( \frac{1}{6} \right)^{\lambda} + 2 \left( \frac{1}{6} \right)^{1 - \lambda} \left( \frac{1}{6} - \varepsilon \right)^{\lambda}
\]

\[
= \frac{1}{3} \times \left( 1 + (1 + 6\varepsilon)^{\lambda} + (1 - 6\varepsilon)^{\lambda} \right)
\]

minimum reached for \( \lambda \approx \frac{1}{2} \)

\[
C(P_0, P_1) \approx -\log(1 - \varepsilon^2) \approx \frac{\varepsilon^2}{\ln 2}
\]

we deduce

\[
\alpha \bullet \beta = 1 - \text{BestAdv}_q \sim e^{-q\varepsilon^2}
\]

For information: the easy bound was \( \text{BestAdv}_q \leq q \times 2\varepsilon \)
Example with Different Supports

given \( a + b = 1 \) s.t. \( \frac{1}{3} > a > \frac{1}{7} \)

\[
P_0 = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \quad P_1 = \begin{pmatrix} a & b & 0 \end{pmatrix}
\]

\[
f(\lambda) = \frac{1}{3} (3a)^\lambda + \frac{1}{3} (3b)^\lambda
\]

we have \( f(0) = \frac{2}{3}, \, f(1) = 1, \, f \) convex, \( f'(0) > 0 \) so

\[
C(P_0, P_1) = -\log \frac{2}{3}
\]

we have

\[
\alpha = \left( \frac{2}{3} \right)^q \quad \beta = (\text{min } f)^q
\]
1 From Statistical Distance to Chernoff Information
   - A Common Cryptographic Problem
   - Hypothesis Testing
   - Best Advantage with Single Sample
   - Chernoff Information
   - Chernoff Bound
     - Approximations of the Chernoff Information
     - Consequence of the Sanov Theorem for Same Support
     - Application to Composite Hypothesis Testing

2 Applications

3 Further Extensions
Let $X_1, \ldots, X_q$ be iid 0-1 random variables with expected value $b$. For $a > b$ we have

\[
\Pr \left[ \frac{1}{q} \sum_{i=1}^{q} X_i \geq a \right] \leq \left( \left( \frac{b}{a} \right)^a \left( \frac{1-b}{1-a} \right)^{1-a} \right)^q = 2^{-qD(a||b)}
\]

Other form:

\[
\sum_{i=[aq]}^{q} \binom{q}{i} b^i (1-b)^{q-i} \leq 2^{-qD(a||b)}
\]
Exercice 1

Consider

\[ P_0 = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ a & 1 - a \end{pmatrix} \quad P_1 = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ b & 1 - b \end{pmatrix} \]

and the distinguisher which outputs 1 iff \( n_1 \leq m \) with \( bq < m < aq \)

1. Show that

\[ 1 - \text{Adv}_q = \sum_{i \leq m} \binom{q}{i} a^i (1 - a)^{q-i} + \sum_{i > m} \binom{q}{i} b^i (1 - b)^{q-i} \]

2. Using the Chernoff bound, show that

\[ 1 - \text{Adv}_q \leq 2^{-qD(\frac{m}{q} \parallel a)} + 2^{-qD(\frac{m}{q} \parallel b)} \]
More General Bound

**Theorem (Chernoff Bound)**

Let $\mathcal{Z}$ be a finite set. Let $P_0$ and $P_1$ be two distributions with support of union $\mathcal{Z}$. Let $\text{BestAdv}_q(P_0, P_1)$ be the best advantage for distinguishing $P_0$ from $P_1$ using $q$ samples. We have

$$1 - \text{BestAdv}_q(P_0, P_1) \leq 2^{-qC(P_0, P_1)}$$
Proof

\[ 1 - \text{BestAdv}_q(P_0, P_1) = \sum_{z^q : \Pr_{P_0}[z^q] > \Pr_{P_1}[z^q]} \Pr_{P_0}[z^q] + \sum_{z^q : \Pr_{P_0}[z^q] < \Pr_{P_1}[z^q]} \Pr_{P_0}[z^q] \]

\[ = \sum_{z^q \in (\text{Supp}(P_0) \cap \text{Supp}(P_1))^q} \min_{P_0, P_1} \left( \Pr[z^q], \Pr[z^q] \right) \]

since \( \min(a, b) \leq a^{1-\lambda} b^\lambda \) for all positive \( a, b \) and \( 0 < \lambda < 1 \) we have

\[ 1 - \text{BestAdv}_q(P_0, P_1) \leq \inf_{0 < \lambda < 1} \sum_{z^q \in (\text{Supp}(P_0) \cap \text{Supp}(P_1))^q} \Pr_{P_0}[z^q]^{1-\lambda} \Pr_{P_1}[z^q]^\lambda \]

\[ = \inf_{0 < \lambda < 1} \sum_{z^q \in (\text{Supp}(P_0) \cap \text{Supp}(P_1))^q} \prod_{i=1}^q P_0(z_i)^{1-\lambda} P_1(z_i)^\lambda \]

\[ = \inf_{0 < \lambda < 1} \left( \sum_{z \in \text{Supp}(P_0) \cap \text{Supp}(P_1)} P_0(z)^{1-\lambda} P_1(z)^\lambda \right)^q \]

\[ = 2^{-qC(P_0, P_1)} \]
1 From Statistical Distance to Chernoff Information
   - A Common Cryptographic Problem
   - Hypothesis Testing
   - Best Advantage with Single Sample
   - Chernoff Information
   - Chernoff Bound
   - Approximations of the Chernoff Information
     - Consequence of the Sanov Theorem for Same Support
     - Application to Composite Hypothesis Testing

2 Applications

3 Further Extensions
Theorem

Let $P_1$ be a variable distribution over $\mathbb{Z}$ which tends towards distribution $P_0$ of support $\mathbb{Z}$. We have

$$C(P_0, P_1) \sim -\log \sum_{x \in \mathbb{Z}} \sqrt{P_0(x)P_1(x)}$$

(the optimal $\lambda$ tends towards $\frac{1}{2}$)

Remark: we always have $C(P_0, P_1) \geq -\log \sum_{x \in \mathbb{Z}} \sqrt{P_0(x)P_1(x)}$
Theorem

Let $P_1$ be a variable distribution over $\mathbb{Z}$ which tends towards distribution $P_0$ of support $\mathbb{Z}$. We have

$$C(P_0, P_1) \sim \frac{1}{8 \ln 2} \sum_{x \in \mathbb{Z}} \frac{(P_1(x) - P_0(x))^2}{P_0(x)}$$

Example for the uniform distribution $P_0$ over $\mathbb{Z}$ of size $n$:

$$C(P_0, P_1) \sim \frac{n}{8 \ln 2} \sum_{x \in \mathbb{Z}} \left(P_1(x) - \frac{1}{n}\right)^2 = \frac{n}{8 \ln 2} \|P_1 - \text{uniform}\|^2_2$$

Squared Euclidean Imbalance (SEI)
Inequality

**Lemma**

Let $P_0$ be a distribution of support $\mathbb{Z}$ and $P_1$ be a distribution over $\mathbb{Z}$. We have

\[
\sum_{x \in \mathbb{Z}} \sqrt{P_0(x)P_1(x)} \leq 1 - 2^{-C(P_0,P_1)} \leq \frac{1}{8} \sum_{x \in \mathbb{Z}} P_0(x) \left( \frac{P_1(x) - P_0(x)}{\min(P_0(x), P_1(x))} \right)^2
\]

Application for $P_0$ uniform over a domain of size $N$: since $P_1(x) \geq P_0(x) - \|P_1 - P_0\|_2$ we have

\[
\frac{1}{\sqrt{N}} \sum_{x \in \mathbb{Z}} \sqrt{P_1(x)} \leq 1 - 2^{-C(P_0,P_1)} \leq \frac{1}{8} \frac{N\|P_1 - P_0\|_2^2}{\left(1 - N\|P_1 - P_0\|_2\right)^2}
\]
Proof — i

Given $0 < \lambda < 1$, let

$$f(\lambda) = \sum_{x \in \mathbb{Z}} P_0(x)^{1-\lambda} P_1(x)^{\lambda}$$

we let $P_1(x) = P_0(x)(1 + \varepsilon_x)$ with $\varepsilon_x \leq \frac{1}{P_0(x)} - 1$

we have

$$f(\lambda) = \sum_{x \in \mathbb{Z}} P_0(x)(1 + \varepsilon_x)^{\lambda}$$

for any $\varepsilon$ we know that

$$(1 + \varepsilon)^{\lambda} - (1 + \lambda \varepsilon) = \frac{\lambda(\lambda - 1)}{2} \varepsilon^2 (1 + \theta \varepsilon)^{\lambda - 2}$$

for some $\theta \in [0, 1]$
Proof — ii

since \( \sum_x P_0(x)(1 + \lambda \varepsilon_x) = 1 \) and \( \sum_x P_0(x)(1 + \varepsilon_x)^\lambda = f(\lambda) \), we obtain

\[
1 - f(\lambda) = \left| \frac{\lambda(\lambda - 1)}{2} \right| \sum_x P_0(x) \varepsilon_x^2 (1 + \theta_x \varepsilon_x)^{\lambda - 2}
\]

\[
= \left| \frac{\lambda(\lambda - 1)}{2} \right| \sum_x P_0(x) \frac{(P_1(x) - P_0(x))^2}{P_0(x)^2} (1 + \theta_x \varepsilon_x)^{\lambda - 2}
\]

if \( \varepsilon_x \geq 0 \) then \( (1 + \theta_x \varepsilon_x)^{\lambda - 2} \leq 1 \) and \( P_0(x) \leq P_1(x) \)

otherwise \( (1 + \theta_x \varepsilon_x)^{\lambda - 2} \leq \frac{P_0(x)^2}{P_1(x)^2} \) and \( P_1(x) \leq P_0(x) \)

finally,

\[
1 - \inf_{0 < \lambda < 1} f(\lambda) \leq \frac{1}{8} \sum_x P_0(x) \left( \frac{P_1(x) - P_0(x)}{\min(P_0(x), P_1(x))} \right)^2
\]
Exercice 2

Consider $b < a$ and

\[ P_0 = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ a & 1 - a \end{pmatrix} \quad P_1 = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ b & 1 - b \end{pmatrix} \]

1. Show that the distinguisher who outputs 1 iff

\[ \frac{n_1}{q} \leq \frac{1}{1 - \frac{\ln b}{\ln \frac{1 - b}{1 - a}}} \]

is a best distinguisher.

2. For $b \to a$, show that this test approximates to

\[ \frac{n_1}{q} \leq \frac{a + b}{2} \]

3. In addition to this, show that

\[ C(P_0, P_1) \sim \frac{(a - b)^2}{8a(1 - a) \ln 2} \]
Exercice 3

Consider

\[
P_0 = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ a & 1-a \end{pmatrix}, \quad P_1 = \begin{pmatrix} 1 & 2 \\ \downarrow & \downarrow \\ b & 1-b \end{pmatrix}
\]

and the distinguisher which outputs 1 iff \( n_1 \leq m \) with \( bq < m < aq \)

1. For \( m = \frac{a+b}{2} q \), compare the concrete expression, the asymptotic expression of the best advantage, and the Chernoff bound for \( 1 - \text{Adv}_q \) for some concrete values for \( a, b, q \)

2. Deduce that the asymptotic expression is pretty good
1 From Statistical Distance to Chernoff Information
   - A Common Cryptographic Problem
   - Hypothesis Testing
   - Best Advantage with Single Sample
   - Chernoff Information
   - Chernoff Bound
   - Approximations of the Chernoff Information
   - Consequence of the Sanov Theorem for Same Support
   - Application to Composite Hypothesis Testing

2 Applications

3 Further Extensions
Theorem

Let \( Z \) be a finite set. Let \( P_0 \) and \( P_1 \) be two distributions with support \( Z \). Let \( \text{BestAdv}_q(P_0, P_1) \) be the best advantage for distinguishing \( P_0 \) from \( P_1 \) using \( q \) samples. We have

\[
1 - \text{BestAdv}_q(P_0, P_1) = 2^{-qC(P_0; P_1)}
\]

Caution: result for \( \alpha \) and \( \beta \) is incorrect without the support assumption.
Lemma

Let $\mathcal{Z}$ be a finite set. Let $P_0$ and $P_1$ be two distributions with support $\mathcal{Z}$. Let

$$\Pi = \left\{ P; \sum_{z \in \mathcal{Z}} P(z) \log \frac{P_0(z)}{P_1(z)} \leq 0 \right\}$$

Let $Y$ be a random vector of $q$ iid samples following $P_0$. We have

$$\Pr[P_Y \in \Pi] \overset{\bullet}{=} 2^{-qC(P_0, P_1)}$$

Proof of the Theorem.

Consider the best distinguisher s.t. output $= 1 \iff P_Y \in \Pi$

$$\alpha = \Pr[P_Y \in \Pi] \overset{\bullet}{=} 2^{-qC(P_0, P_1)}$$

By exchanging $P_0$ and $P_1$ we obtain $\Pr[P_Y \not\in \Pi] \overset{\circ}{=} 2^{-qC(P_0, P_1)}$

Since $\beta \leq \Pr[P_Y \not\in \Pi] \overset{\circ}{=} 2^{-qC(P_0, P_1)}$ we have

$$1 - \text{BestAdv}_q(P_0, P_1) = \alpha + \beta \overset{\bullet}{=} 2^{-qC(P_0, P_1)}$$
Proof of Lemma — i

- use Sanov: we just have to prove $\bar{\Pi} = \bar{\Pi}$ and $D(\Pi \| P_0) = C(P_0, P_1)$
- wlog $P_0 \neq P_1$ so there exists $x \in \mathbb{Z}$ s.t. $0 < P_0(x) < P_1(x)$
  - clearly, the Dirac distribution on point $x$ is in $\bar{\Pi}$ so $\bar{\Pi}$ is not empty and contains an open ball
- $\Pi$ is convex and $\bar{\Pi}$ is non empty so $\bar{\Pi} = \bar{\Pi}$ hence $\Pr[P_Y \in \Pi] \geq 2^{-qD(\Pi \| P_0)}$
- $\Pi$ is topologically closed in a compact space hence it is compact
  - $P \mapsto D(P \| P_0)$ is continuous
  - hence, there must exist some $P$ s.t. $D(P \| P_0) = D(\Pi \| P_0)$
- since $P \mapsto -D(P \| P_0)$ is convex, a local minimum must be global
  - any segment from a local minimum to $P_0$ must be outside $\Pi$
- conclusion: $D(\Pi \| P_0) = D(P \| P_0)$ for a local minimum $P$ verifying $\sum_z P(z) \log \frac{P_0(z)}{P_1(z)} = 0$
Proof of Lemma — ii

- using the Lagrange multiplyers, \( P \) must satisfy

\[
\frac{\partial D(P \| P_0)}{\partial P(z)} = \alpha + \beta \log \frac{P_0(z)}{P_1(z)}
\]

for all \( z \), with some constant \( \alpha \) and \( \beta \)

we deduce \( P = P_\lambda \) for some \( \lambda \) where

\[
P_\lambda(z) = \frac{P_0(z)^{1-\lambda}P_1(z)^{\lambda}}{\sum_a P_0(a)^{1-\lambda}P_1(a)^{\lambda}}
\]

- let \( f(\lambda) = \log \sum_a P_0(a)^{1-\lambda}P_1(a)^{\lambda} \)

we have \( D(P_\lambda \| P_0) = -\lambda L(P_\lambda) - f(\lambda) \) and \( f'(\lambda) = -L(P_\lambda) \)

where \( L(P) \leq 0 \) defines \( \Pi \)

- we have \( f'(0) = -D(P_0 \| P_1) < 0 \) and \( f(0) = f(1) = 0 \) so there exists \( \lambda \in ]0, 1[ \) such that \( f'(\lambda) = 0 \) and for which \( f(\lambda) \) is minimal for this \( \lambda \), we deduce \( L(P_\lambda) = 0 \) and

\[
D(P_\lambda \| P_0) = -f(\lambda) = -\inf_{]0,1[} f = C(P_0, P_1)
\]
1 From Statistical Distance to Chernoff Information
   - A Common Cryptographic Problem
   - Hypothesis Testing
   - Best Advantage with Single Sample
   - Chernoff Information
   - Chernoff Bound
   - Approximations of the Chernoff Information
   - Consequence of the Sanov Theorem for Same Support
   - Application to Composite Hypothesis Testing

2 Applications

3 Further Extensions
Composite Alternate Hypothesis

**Hypothesis** $H_0$: variables follow distribution $P_0$

**Hypothesis** $H_1$: variables follow one of the distributions $P_1, \ldots, P_d$

we assume that $P_0, \ldots, P_d$ are known

idea: use the likelihood ratio with $P_0$ against the $P_i$ which is the closest (to the sense of $D(\cdot \| P_i)$):

\[
\Pi = \left\{ P; \min_{1 \leq i \leq d} \sum_{x \in \mathbb{Z}} P(x) \log \frac{P_0(x)}{P_i(x)} \leq 0 \right\}
\]

\[
= \left\{ P; \min_{1 \leq i \leq d} D(P \| P_i) \leq D(P \| P_0) \right\}
\]
Result

Hypothesis $H_0$: variables follow distribution $P_0$

Hypothesis $H_1$: variables follow one of the distributions $P_1, \ldots, P_d$

Theorem

The best distinguisher satisfies

$$1 - \text{Adv} \equiv \max_{1 \leq i \leq d} 2^{-qC(P_0, P_i)}$$

for any (nonzero) weights on the $P_i$'s.

input: $x_1, \ldots, x_q$

threshold: $\tau$

1: $L = \min_{1 \leq i \leq d} \sum_{j=1}^{q} \log \frac{P_0(x_j)}{P_i(x_j)}$

2: if $L \leq \log \tau$ then

3: $b \leftarrow 1$

4: else

5: $b \leftarrow 0$

6: end if

output: $b$

we need a number of samples $\sim 1 / \min_i C(P_0, P_i)$
Proof — i

We assume that $P_i$ under $H_1$ is selected with probability $\pi_i \neq 0$.

- for any distinguisher limited to $q$ queries we have

\[
1 - \text{Adv}(H_0, H_1) = \sum_{i=1}^{d} \pi_i (1 - \text{Adv}(P_0, P_i))
\]

but since $1 - \text{Adv}(P_0, P_i) \geq 1 - \text{BestAdv}(P_0, P_i) = 2^{-qC(P_0,P_i)}$ we have

\[
1 - \text{Adv}(H_0, H_1) > \sum_{i=1}^{d} \pi_i 2^{-qC(P_0,P_i)} = \max_{1 \leq i \leq d} 2^{-qC(P_0,P_i)}
\]
Proof — ii

Let $L = \min L_i$ where $L_i$ is the_llr for distinguishing $P_0$ from $P_i$. Let $\Pi$ resp. $\Pi_i$ is the set of all type s.t. $L \leq 0$ resp. $L_i \leq 0$ so $\Pi = \bigcup_i \Pi_i$. If the $x$’s follow $P_i$ then $\Pr[\text{output } 0|P_i]$ is the probability that all $L_j$’s including $L_i$ are positive. So $\Pr[\text{output } 0|P_i] \leq \Pr[L_i \geq 0]$ which is the error type II for the best distinguisher between $P_0$ and $P_i$ which is $2^{-qC(P_0,P_i)}$. So

$$\Pr[\text{output } 0|H_1] \cdot \sum_i \pi_i 2^{-qC(P_0,P_i)} \cdot \max_i 2^{-qC(P_0,P_i)}$$

If the $x$’s follow $P_0$ then, thanks to Sanov

$$\Pr[\text{output } 1|H_0] \cdot 2^{-qD(\Pi||P_0)} = \max_{1 \leq i \leq d} 2^{-qD(\Pi_i||P_0)} \cdot \max_{1 \leq i \leq d} 2^{-qC(P_0,P_i)}$$

Finally,

$$1 - \text{Adv}(H_0, H_1) = \Pr[0|H_1] + \Pr[1|H_0] \cdot \max_i 2^{-qC(P_0,P_i)}$$
Exercice 4

Detail the best distinguisher and estimate the advantage for

**Hypothesis** $H_0$: variables follow distribution $P_0$

**Hypothesis** $H_1$: variables follow $P$ such that $\|P_0 - P\|_r = d$

for

- $r = 2$
- $r = \infty$
- $r = 1$

Reminder:

$$\|f\|_2 = \sqrt{\sum_x |f(x)|^2} , \quad \|f\|_\infty = \max_x |f(x)| , \quad \|f\|_1 = \sum_x |f(x)|$$
1. From Statistical Distance to Chernoff Information
2. Applications
3. Further Extensions
1. From Statistical Distance to Chernoff Information

2. Applications
   - Application to Block Cipher Analysis
   - The Leftover Hash Lemma
   - Soundness Amplification
   - CAPTCHA-Like Challenge-Response Protocols

3. Further Extensions
Distinguishing Attack on Ciphers

Indistinguishability from an ideal scheme is another security model

- $C$: permutation (block cipher) defined by a random key
- $C^*$: uniformly distributed random permutation (ideal scheme)
- Advantage: $\Pr[\text{output} = 1 | C] - \Pr[\text{output} = 1 | C^*]$
Applying the Theory about Distinguishing Sources

- \( x = (x_1, \ldots, x_d) \), \( y = (y_1, \ldots, y_d) \), \( y_i = c(x_i) \) for all \( i \)

- assume that

  if \( x \) follows distribution \( D \) and \( c \) is a fixed \( C \) resp \( C^* \) then

  \[ z = h(x, y) \] follows distribution \( P \) resp \( P^* \)

Examples:

- differential cryptanalysis (\( d = 2 \)): \( x_1 \) random and \( x_2 = x_1 \oplus a \) then

  \( y_1 \oplus y_2 \) biased (for \( C \)) or uniform (for \( C^* \))

- linear cryptanalysis (\( d = 1 \)): \( x \) random then \( (a \cdot x) \oplus (b \cdot y) \) biased

  (for \( C \)) or a fair coin (for \( C^* \))

- defined by \( d, D \) and \( h \), a distinguisher between \( P \) and \( P^* \) defines an **iterative distinguisher** between \( C \) and \( C^* \)
Iterative Distinguisher

iterative distinguisher of order $d$:

**Parameters:** a complexity $q$

**Oracle:** a permutation $c$

1. for $i$ from 1 to $q$ do
2. \hspace{1em} pick $X_1, \ldots, X_d$ following $D$
3. \hspace{1em} query for $Y_j = c(X_j), j = 1, \ldots, d$
4. \hspace{1em} set $Z_i = h(X_1, \ldots, X_d, Y_1, \ldots, Y_d)$
5. end for
6. apply optimal distinguisher on $Z_1, \ldots, Z_q$

we say this is an optimal $h$-distinguisher
Example: Differential Distinguisher — i

- $h(x, y) = y_1 \oplus y_2$, $P^*$ uniform
- we assume $P(b) = p$ such that $\frac{1}{n} = o(p)$ and $p = o(1)$, and $P(c) = \frac{1-p}{n-1} = \beta$ for each $c \neq b$
- let

$$f(\lambda) = \sum_c P(c)^{1-\lambda} \frac{1}{n^{\lambda}} = \frac{p}{(np)^{\lambda}} + \frac{1-p}{(n\beta)^{\lambda}}$$

we have $f(0) = f(1) = 1$ and $f'(0) \leq 0$ so we define $\lambda_0$ such that $f'(\lambda_0) = 0$ and get

$$\lambda_0 = \ln \frac{p \ln(np)}{(1-p) \ln \frac{1-1/n}{1-p}} \sim \frac{\ln \ln(np)}{\ln(np)}$$

so $(np)^{\lambda_0} \sim \ln(np)$ and $(n\beta)^{\lambda_0} = 1 + o(p)$ thus $f(\lambda_0) = 1 - p + o(p)$ therefore the Chernoff information is

$$C(P, P^*) = - \log f(\lambda_0) \sim \frac{p}{\ln 2}$$

so we need $q \approx \ln 2/p$ samples to run the best distinguisher
Example: Differential Distinguisher — ii

- likelihood ratio is \( R \approx \left( \frac{1}{(np)^{n_b}(n\beta)^{q-n_b}} \right) \)
- best distinguisher yields 1 iff \( R \leq 1 \) which is equivalent to

\[
\frac{n_b}{q} \geq \frac{\ln(n\beta)}{\ln(\beta/p)} \sim \frac{p}{\ln(np)}
\]

since we take \( q \approx \ln 2/p \) this condition is equivalent to \( n_b > 0 \)
Example: Differential Distinguisher — iii

Parameters: a complexity $q$
Oracle: a permutation $c$

1. for $i$ from 1 to $q$ do
2. pick uniformly a random $X$
3. query for $c(X)$ and $c(X \oplus a)$
4. if $c(X \oplus a) = c(X) \oplus b$, output 1 and stop
5. end for
6. output 0

Theorem

If $\frac{1}{n} \ll p \ll 1$, the following hypothesis testing problem is performed with a significant advantage using $q \approx \ln 2/p$ samples.

Hypothesis $H_0$: $\text{DP}^c(a, b) = \frac{1}{n}$
Hypothesis $H_1$: $\text{DP}^c(a, b) = p$
Example: Impossible Differential

Same with $p = 0$:

- we have $f(\lambda) = \left(1 - \frac{1}{n}\right)^\lambda$
- so, $C(P, P^*) = -\log \left(1 - \frac{1}{n}\right) \sim \frac{1}{n \ln 2}$
- we thus need $q \approx n \ln 2$
- best distinguisher yields 1 iff $n_b = 0$
Consider the treatment on differential distinguishers.
With the same function $h$ but assumption $p = o(1/n)$ instead of $\frac{1}{n} \ll p$, recompute $C(P, P^*)$ and obtain the data complexity of the improbable differential distinguisher.
Example: Linear Distinguisher — i

see previous computation

- \( h(x, y) = (a \cdot x) \oplus (b \cdot y) \), \( P^* \) uniform
- we assume \( P(z) = \frac{1}{2}(1 \pm \varepsilon) \) such that \( \varepsilon = o(1) \) for all \( z \)
  (composite hypothesis)
- we have

\[
C(P, P^*) \approx \frac{2}{8 \ln 2} \sum_{z=0}^{1} \left( P(z) - \frac{1}{2} \right)^2 = \frac{\varepsilon^2}{8 \ln 2}
\]

so we need \( q \approx \frac{8 \ln 2}{\varepsilon^2} \) samples to run the best distinguisher

- best distinguisher yields 1 iff

\[
\left| \frac{2n_0}{q} - 1 \right| \geq \frac{|\varepsilon|}{2}
\]
Example: Linear Distinguisher — ii

**Parameters:** a complexity $q$

**Oracle:** a permutation $c$

1: initialize the counter value $m$ to zero
2: **for** $i$ from 1 to $q$ **do**
3: pick uniformly a random $X$
4: query for $c(X)$
5: if $a \cdot X = b \cdot c(X)$, increment the counter $m$
6: **end for**
7: output 1 iff $\left| 2 \frac{m}{q} - 1 \right| \geq \frac{|\varepsilon|}{2}$

**Theorem**

If $\varepsilon \ll 1$, the following hypothesis testing problem is performed with a significant advantage using $q \approx 8 \ln 2 / \varepsilon^2$ samples.

**Hypothesis $H_0$:** $\text{LP}^c(a, b) = 0$

**Hypothesis $H_1$:** $\text{LP}^c(a, b) = \varepsilon^2$
From Statistical Distance to Chernoff Information

Applications
- Application to Block Cipher Analysis
- The Leftover Hash Lemma
- Soundness Amplification
- CAPTCHA-Like Challenge-Response Protocols

Further Extensions
Definitions

- **min-entropy:**

  \[ H_\infty(X) = -\log \max_x \Pr[X = x] \]

- **universal hash function:**

  \[
  \forall x \neq x' \quad \Pr[N(h_N(x) = h_N(x'))] = \frac{1}{\#\text{range}}
  \]

  where range is the output domain of \( h \) and \( N \) is uniformly distributed

- **Rényi entropy:**

  \[ H_2(X) = -\log \sum_x \Pr[X = x]^2 \]

  \( 2^{-H_2(X)} \) is the **collision probability**
Euclidean distance and Rényi entropy:

\[ \| \text{distr}(X) - \text{uniform} \|_2^2 = 2^{-H_2(X)} - \frac{1}{\#\text{domain}} \]

Rényi entropy and min-entropy

\[ 2^{-H_2(X)} \leq 2^{-H_\infty(X)} \]

Statistical distance and Euclidean distance:

\[ d(\text{distr}(X), \text{uniform}) \leq \| \text{distr}(X) - \text{uniform} \|_2 \sqrt{\#\text{domain}} \]
Leftover Hash Lemma

Lemma (Impagliazzo-Levin-Luby 1989)

If \( m \leq H_\infty(X) - 2 \log \frac{1}{\varepsilon} \) and \( h \) is a universal hash function with a range of size \( 2^m \) then \((h_N(X), N)\) and \((U, N)\) have distributions which are \( \varepsilon \)-indistinguishable.

\( X, N, U \) are independent.
\( N \) and \( U \) are uniformly distributed.
Proof

We denote $P_0$ and $P_1$ the distributions and compute the Euclidean distance:

$$
\| P_1 - P_0 \|_2^2 = \sum_{k,n} \left( \Pr_{X,N}[h_n(X) = k, N = n] - \frac{1}{2^m \# \mathcal{N}} \right)^2
$$

$$
= \frac{1}{(\# \mathcal{N})^2} \sum_{k,n} \Pr_{X,X'}[h_n(X) = h_n(X') = k] - \frac{1}{2^m \# \mathcal{N}}
$$

$$
= \frac{1}{\# \mathcal{N}} \sum_{x,x'} \Pr[X = x, X' = x', h_N(x) = h_N(x')] - \frac{1}{2^m \# \mathcal{N}}
$$

$$
= \frac{1 - 2^{-m}}{\# \mathcal{N}} \sum_x \Pr[X = x]^2
$$

$$
\leq \frac{1 - 2^{-m}}{\# \mathcal{N}} 2^{-H_\infty(X)} \leq \frac{1}{2^m \# \mathcal{N}} \varepsilon^2
$$

we then use the link between statistical distance and Euclidean distance to obtain $d(P_0, P_1) \leq \varepsilon$
Assume a subgroup $\langle g \rangle$ generated by some $g$ of prime order $q$ in $\mathbb{Z}_p^*$

$$K_S = x \in \mathbb{Z}_q^* \quad \text{Enc}(K_P, m; r) = (g^r, my^r) \quad r \in \mathbb{Z}_q^*$$

$$K_P = g^x \quad \text{Dec}(K_S, u, v) = vu^{-x}$$

- key recovery is equivalent to the discrete logarithm problem
- decryption is equivalent to the Diffie-Hellman problem
- not semantically secure:
  - $g^{p-1 \over 2} = 1$ since $q$ must divide $p-1 \over 2$
  - thus $(g/p) = +1$
  - we deduce $(my^r/p) = (m/p)$
  - if $(m_b/p) = (-1)^b$ for $b = 0, 1$ we can distinguish $\text{Enc}(K_P, m_0; r)$ from $\text{Enc}(K_P, m_1; r)$ with advantage 1
Application: ElGamal Encryption — ii

Assume a group \( \langle g \rangle \) generated by some \( g \) of prime order \( q \)

\[
K_S = x \in \mathbb{Z}_q^* \quad \text{Enc}(K_P, m; N, r) = (g^r, m \oplus h_N(y^r), N) \quad r \in \mathbb{Z}_q^* \\
K_P = g^x \quad \text{Dec}(K_S, u, v, N) = v \oplus h_N(u^x)
\]

- due to the DDH assumption, \( (g, g^r, m \oplus h_N(y^r), N) \) is \( \varepsilon_{\text{DDH}} \)-indistinguishable from \( (g, g^r, m \oplus h_N(g^r'), N) \)
- due to Lemma, \( (g, g^r, m \oplus h_N(g^r'), N) \) is \( \varepsilon \)-indistinguishable from \( (g, g^r, m \oplus U, N) \)
- \( (g, g^r, m \oplus U, N) \) is perfectly indistinguishable from \( (g, g^r, U, N) \)
- consequently, \( (g, g^r, m \oplus h_N(y^r), N) \) is \( (\varepsilon_{\text{DDH}} + \varepsilon) \)-indistinguishable from something independent from \( m \)
- so the scheme is \( (\varepsilon_{\text{DDH}} + \varepsilon) \)-IND-CPA
Application: Diffie-Hellman with Key Derivation

Assume a group $\langle g \rangle$ generated by some $g$ of prime order $q$

Alice

- pick $x \in \mathbb{Z}_q^*$, $X \leftarrow g^x$
- if $Y \not\in \langle g \rangle - \{1\}$, abort
- $K \leftarrow Y^x$

Bob

- $X \leftarrow g^x$
- if $X \not\in \langle g \rangle - \{1\}$, abort
- pick $y \in \mathbb{Z}_q^*$, $Y \leftarrow g^y$
- $K \leftarrow X^y$

$(K_{\text{raw}} = g^{xy})$

since $\mathbb{Z}_q^*$ is cyclic, $K_{\text{raw}}$ is a uniformly distributed non-neutral element of $\langle g \rangle$ (even locally under active attack)
Key Derivation

- assume a non-ambiguous representation format for values which may be in \(\langle g \rangle\) or not
- \(\Pr[K_{\text{raw}} = x] = 0\) or \((q - 1)^{-1}\) for all value \(x\)

\[H_\infty(K_{\text{raw}}) = \log(q - 1)\]

- exchange a random number \(N\) and derive the key \(K = h_N(K_{\text{raw}})\)
  \(\rightarrow\) indistinguishable from a random key

- a protocol using \(n\) such key generations is \(n\varepsilon\)-indifferentiable from the same protocol where \(K\) is truly random

- this comes from the trivial bound: we could do better
Multi-Sample Leftover Hash Lemma

Lemma

If \( m \leq H_\infty(X) - 2\log \frac{1}{\varepsilon} \) and \( m \leq \frac{H_\infty(X) + \#\mathcal{N}}{2} - \log \frac{1}{\varepsilon'} \) and \( h \) is a universal hash function with a range of size \( 2^m \) and key space \( \mathcal{N} \) then \((h_N(X), N)\) and \((U, N)\) have distributions such that for any distinguisher using \( q \) queries we have

\[
\text{BestAdv}_q \leq (q + o(q)) \frac{\varepsilon^2}{8(1 - \varepsilon')^2}
\]

\( X, N, U \) are independent.
\( N \) and \( U \) are uniformly distributed.

we get \( q\frac{\varepsilon^2}{8} \) instead of \( q\varepsilon \)
Proof

We denote \( P_0 \) and \( P_1 \) the distributions

- we already have proven \( \| P_1 - P_0 \|_2^2 \leq 2^{-H_\infty(X)/\#\mathcal{N}} \) and the domain is of size \( 2^m \#\mathcal{N} \)
- using the upper bound on \( 1 - 2^{-C(P_0, P_1)} \) we get

\[
1 - 2^{-C(P_0, P_1)} \leq \frac{2^{m-H_\infty(X)}}{8 \left( 1 - \sqrt{2^{2m-H_\infty(X)} \#\mathcal{N}} \right)^2}
\]

\[
\leq \frac{\varepsilon^2}{8 \left( 1 - \varepsilon' \right)^2}
\]

- hence

\[
\text{BestAdv}_q \leq 1 - e^{o(q)} \left( 1 - 2^{-C(P_0, P_1)} \right)^q \leq \left( q + o(q) \right) \frac{\varepsilon^2}{8(1-\varepsilon')^2}
\]
From Statistical Distance to Chernoff Information

Applications
- Application to Block Cipher Analysis
- The Leftover Hash Lemma
- Soundness Amplification
- CAPTCHA-Like Challenge-Response Protocols

Further Extensions
Interactive Proof

**Definition**

Given a language $L$ over an alphabet $Z$, an **interactive proof system** for $L$ is a pair $(P, V)$ of interactive machines such that there exists a polynomial $P$, $a$, $b$ such that $0 \leq b < a \leq 1$ and

- **termination**: for any $x$, the total complexity of $V$ (until termination) in $P \leftrightarrow V(r)$ is bounded by $P(|x|)$
- **$a$-completeness**: for any $x \in L$ then

\[
\Pr_{r_P, r_V} \left[ \text{Out}_V \left( P(r_P) \leftrightarrow V(r_V) \right) = \text{accept} \right] \geq a
\]

- **$b$-soundness**: for any $x \notin L$ and any algorithm $P^*$ then

\[
\Pr_{r_P, r_V} \left[ \text{Out}_V \left( P^*(r_P) \leftrightarrow V(r_V) \right) = \text{accept} \right] \leq b
\]
Sequential Composition — i

Given an interactive proof system \((\mathcal{P}, \mathcal{V})\) for \(L\) which is \(a\)-complete and \(b\)-sound we define an new proof system \((\mathcal{P}', \mathcal{V}')\) as follows:

- \(\mathcal{P}'\) resp \(\mathcal{V}'\) simulates \(\mathcal{P}\) resp \(\mathcal{V}\) but have no terminal message until \(q\) iterations are made
- after an iteration completes, they restart the entire protocol with fresh random coins
- \(\mathcal{V}'\) accepts if at least \(m\) iterations accepted out of \(q\)

the new interactive proof system is \(a'\)-complete and \(b'\)-sound with

\[
a' = \sum_{i=m}^{q} \binom{q}{i} a^i (1 - a)^{q-i}
\]

\[
b' = \sum_{i=m}^{q} \binom{q}{i} b^i (1 - b)^{q-i}
\]
Sequential Composition — ii

\[
a' = \sum_{i=m}^{q} \binom{q}{i} a^i (1 - a)^{q-i}
\]

\[
b' = \sum_{i=m}^{q} \binom{q}{i} b^i (1 - b)^{q-i}
\]

by taking \( m = \frac{q}{(1 - \ln(b/a)/\ln((1 - b)/(1 - a)))} \) we have

\[
a' - b' = \text{BestAdv}_q(a, b) = 1 - 2^{-qC(a,b) + o(q)}
\]

so \( a' \to 1 \) and \( b' \to 0 \) exponentially fast
problem: adversaries are not forced to independently treat each iteration

how to prove

\[ b' \leq \sum_{i=m}^{q} \binom{q}{i} b^i (1 - b)^{q-i} \]
From Statistical Distance to Chernoff Information

Applications
- Application to Block Cipher Analysis
- The Leftover Hash Lemma
- Soundness Amplification
- CAPTCHA-Like Challenge-Response Protocols

Further Extensions
Problem Statement

Impagliazzo-Jaiswal-Kabanets CRYPTO 2007

- we have a fuzzy challenge-response protocol
e.g. CAPTCHA
- honest people pass with probability $a$
- malicious people pass with probability $b$
- what is the best way to distinguish using $q$ attempts?

- translation: accept is a 1-0 random variable

**Hypothesis $H_0$:** $E(accept) = a$

**Hypothesis $H_1$:** $E(accept) = b$
Theorem (Impagliazzo-Jaiswal-Kabanets 2007)

If all malicious algorithms pass with probability at most $b$ then the probability that a malicious algorithm passes at least $m$ out of $q$ is lower than $\beta = 2e^{-\frac{(m-bq)^2}{64q}}$

Application:

- replace “pass” by “fail”, $b$ by $1 - a$, and $m$ by $q - m$
- get that honest people succeed less than $m$ out of $q$ with probability lower than $\alpha = 2e^{-\frac{(m-aq)^2}{64q}}$
- so, the advantage using threshold $m$ is

$$1 - \text{Adv}_q \leq \alpha + \beta = 2e^{-\frac{(m-bq)^2}{64q}} + 2e^{-\frac{(m-aq)^2}{64q}}$$
Theorem (Impagliazzo-Jaiswal-Kabanets 2009)

If all malicious algorithms pass with probability at most \( b \) then the probability that a malicious algorithm passes at least \( m \) out of \( q \) is lower than

\[
\beta = \frac{100q}{m-bq} e^{-\frac{(m-bq)^2}{40q(1-b)}}
\]

Application:

\[
1 - \text{Adv}_q \leq \frac{100q}{m-bq} e^{-\frac{(m-bq)^2}{40q(1-b)}} + \frac{100q}{aq-m} e^{-\frac{(m-aq)^2}{40qa}}
\]
Optimization

- Using our General Treatment and (general) Chernoff bound:

\[ 1 - \text{BestAdv}_q \leq 2^{-qC(a,b)} \approx e^{-q \frac{(a-b)^2}{8a(1-a)}} \quad \text{if } a \approx b \]

- Concretely: using Exercise, the best distinguisher yields 1 iff \( n_{\text{success}} \leq m \)

\[ m = \frac{q}{1 - \frac{\log \frac{b}{a}}{\log \frac{1-b}{1-a}}} \approx q \frac{a + b}{2} \quad \text{if } a \approx b \]

\[ 1 - \text{BestAdv}_q = \sum_{i \leq m} \binom{q}{i} a^i (1 - a)^{q-i} + \sum_{i > m} \binom{q}{i} b^i (1 - b)^{q-i} \]

- With Chernoff bound:

\[ 1 - \text{Adv}_q \leq 2^{-qD\left(\frac{m}{q} \| a\right)} + 2^{-qD\left(\frac{m}{q} \| b\right)} \]
Application

\[ IJK07 \quad 1 - \text{Adv}_q \leq 2e^{-\frac{(m-bq)^2}{64q}} + 2e^{-\frac{(m-aq)^2}{64q}} \]

\[ IJK09 \quad 1 - \text{Adv}_q \leq \frac{100q}{m-bq} e^{-\frac{(m-bq)^2}{40q(1-b)}} + \frac{100q}{aq-m} e^{-\frac{(m-aq)^2}{40qa}} \]

asympt. \quad 1 - \text{BestAdv}_q \leq 2^{-qC(a,b)}

concrete \quad 1 - \text{Adv}_q = 1 - \sum_{i \leq m} \binom{q}{i} (b^i(1-b)^{q-i} - a^i(1-a)^{q-i})

Chernoff \quad 1 - \text{BestAdv}_q \leq 2^{-qD\left(\frac{m}{q} \parallel a\right)} + 2^{-qD\left(\frac{m}{q} \parallel b\right)}

Application to \( a = 90\% \) and \( b = 33\% \): \( 1/C(a, b) = 3.156 \):

<table>
<thead>
<tr>
<th>( q )</th>
<th>( m )</th>
<th>( IJK07 )</th>
<th>( IJK09 )</th>
<th>asympt.</th>
<th>concrete</th>
<th>Chernoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>( &gt;1 )</td>
<td>n/a</td>
<td>0.803</td>
<td>0.430</td>
<td>1.606</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>( &gt;1 )</td>
<td>( &gt;1 )</td>
<td>0.517</td>
<td>0.283</td>
<td>1.035</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>( &gt;1 )</td>
<td>( &gt;1 )</td>
<td>0.415</td>
<td>0.160</td>
<td>0.831</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>( &gt;1 )</td>
<td>( &gt;1 )</td>
<td>0.333</td>
<td>0.125</td>
<td>0.667</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>( &gt;1 )</td>
<td>( &gt;1 )</td>
<td>0.215</td>
<td>0.069</td>
<td>0.430</td>
</tr>
<tr>
<td>100</td>
<td>65</td>
<td>( &gt;1 )</td>
<td>( &gt;1 )</td>
<td>( 2^{-31.7} )</td>
<td>( 2^{-34.9} )</td>
<td>( 2^{-30.7} )</td>
</tr>
<tr>
<td>5000</td>
<td>3273</td>
<td>0.019</td>
<td>0.095</td>
<td>( \approx 0 )</td>
<td>( \approx 0 )</td>
<td>( \approx 0 )</td>
</tr>
</tbody>
</table>
From Statistical Distance to Chernoff Information

Applications

Further Extensions
1 From Statistical Distance to Chernoff Information

2 Applications

3 Further Extensions
   - Spectral Analysis
   - Seek for the Best Efficient Distinguisher
   - Example of Extended Linear Cryptanalysis
Definition

Given a finite Abelian group \( G \):
- A character \( \chi \) is a group homomorphism from \( G \) to \( \mathbb{C}^* \)
- The dual group \( \hat{G} \) is the group of characters over \( G \).

Fact: \( \hat{G} \) is isomorphic to \( G \).

Example: in \( \mathbb{Z}_m^\ell \), for each character \( \chi \) there exists \( u \in \mathbb{Z}_m^\ell \) such that for all \( x \) we have

\[
\chi(x_1, \ldots, x_\ell) = e^{\frac{2i\pi}{m}(u_1x_1 + \cdots + u_\ell x_\ell)}
\]

Example for \( m = 2 \):

\[
\chi(x_1, \ldots, x_\ell) = (-1)^{u_1x_1 + \cdots + u_\ell x_\ell}
\]
Orthogonality of Characters

\[ \langle f, g \rangle = \sum_{x \in G} f(x) \overline{g}(x) \]

**Theorem**

For \( \chi_1, \chi_2 \in \hat{G} \) we have

\[ \langle \chi_1, \chi_2 \rangle = \begin{cases} \#G & \text{if } \chi_1 = \chi_2 \\ 0 & \text{otherwise} \end{cases} \]

The characters form an orthogonal basis of functions from \( G \) to \( \mathbb{C} \).

\[ \hat{f}(\chi) = \langle f, \chi \rangle \]

\[ f(x) = \frac{1}{\#G} \sum_{\chi \in \hat{G}} \hat{f}(\chi) \chi(x) \]

\[ \langle f, f \rangle = \sum_{x \in G} |f(x)|^2 = \frac{1}{\#G} \sum_{\chi \in \hat{G}} |\hat{f}(\chi)|^2 = \frac{1}{\#G} \langle \hat{f}, \hat{f} \rangle \]

this is the discrete Fourier transform
Example based on \( \mathbb{Z}_2 \)

- let \( G = \mathbb{Z}_2^\ell \)
- all characters \( \chi \) are defined by some \( u \) by \( \chi(x) = (-1)^{u \cdot x} \)
- Fourier transform:

\[
\hat{f}(u) = \langle f, u \rangle = \sum_{x \in \mathbb{Z}_2^\ell} f(x)(-1)^{u \cdot x}
\]

- inversion:

\[
f(x) = 2^{-\ell} \sum_{u \in \mathbb{Z}_2^\ell} \hat{f}(\chi)(-1)^{u \cdot x}
\]

- Parseval:

\[
\langle f, f \rangle = \sum_{x \in \mathbb{Z}_2^\ell} |f(x)|^2 = 2^{-\ell} \sum_{u \in \mathbb{Z}_2^\ell} |\hat{f}(u)|^2 = 2^{-\ell} \langle \hat{f}, \hat{f} \rangle
\]
Definition

Given a finite Abelian group $G$, a character $\chi$ over $G$ and a random variable $X$ of distribution $P$, we define

$$LP(\chi(X)) = |E(\chi(X))|^2 = \left| \sum_{x \in G} P(x)\chi(x) \right|^2 = |\hat{P}(\chi)|^2$$

for $G = Z_2^\ell$,

$$LP(\chi_u(X)) = \left| E \left( (-1)^{u \cdot X} \right) \right|^2 = (2 \Pr[u \cdot X = 0] - 1)^2$$
Link with SEI

for all distribution $P$ and the uniform distribution $U$:

- $\hat{P}(1) = \langle P, 1 \rangle = \sum_x P(x) = 1$ hence $\hat{P}(1) = \hat{U}(1)$
- for all $\chi \neq 1$ we have $\hat{U}(\chi) = \langle U, \chi \rangle = \frac{1}{n} \sum_x \chi(x) = 0$
- SEI:

$$SEI(X) = \#G \times \sum_x \left( P(x) - \frac{1}{\#G} \right)^2$$

$$= \sum_{\chi} \left| \hat{P}(\chi) - \hat{U}(\chi) \right|^2$$

$$= \sum_{\chi \neq 1} |\hat{P}(\chi)|^2$$

$$= \sum_{\chi \neq 1} \text{LP}(\chi(X))$$

- hence, if $P$ is close to $U$,

$$C(P, U) \sim \frac{1}{8 \ln 2} \sum_{\chi \neq 1} \text{LP}(\chi(X))$$

hence, if $P$ is close to $U$, 

$$C(P, U) \sim \frac{1}{8 \ln 2} \sum_{\chi \neq 1} \text{LP}(\chi(X))$$
If $X$ is a random variable over a group $G$ with distribution $P$ which tends towards the uniform distribution $U$ we have:

\[ 1 - \text{BestAdv}_q(P, U) \approx 2^{-qC(P, U)} \]

\[ C(P, U) \approx \frac{1}{8 \ln 2} \text{SEI}(P) \]

\[ \text{SEI}(P) = \#G \times \sum_{x \in G} \left( P(x) - \frac{1}{\#G} \right)^2 = \sum_{\chi \neq 1} \text{LP}(\chi(X)) \]

\[ \text{LP}(\chi(X)) = |E(\chi(X))|^2 = |\hat{P}(\chi)|^2 \]

\[ 1 - \text{BestAdv}_q(P, U) \approx \exp \left( -\frac{q}{8} \sum_{\chi \neq 1} \text{LP}(\chi(X)) \right) \]
1. From Statistical Distance to Chernoff Information

2. Applications

3. Further Extensions
   - Spectral Analysis
   - Seek for the Best Efficient Distinguisher
   - Example of Extended Linear Cryptanalysis
**Best Distinguisher**

**input:** \( x_1, \ldots, x_q \)

**threshold:** \( \tau = 1 \)

1. \( L = \sum_{i=1}^{q} \log \frac{P_0(x_i)}{P_1(x_i)} \)
2. if \( L \leq \log \tau \) then
3. \( b \leftarrow 1 \)
4. else
5. \( b \leftarrow 0 \)
6. end if

**output:** \( b \)

**problem:** must know all \( P_0(x_i)/P_1(x_i) \), hard if support is huge
Projection-Based Distinguisher

problem: what is the best way to hash?
Using Characters as Projection

- Let $\chi \in \hat{\mathbb{Z}}$ of order $d$.

  $\chi : \mathbb{Z} \rightarrow \left\{ e^{\frac{2i\pi}{d}j} ; j \in \mathbb{Z}_d \right\}$

  $x \mapsto \chi(x)$

- Let $S$ be the subset of $\hat{\mathbb{Z}}$ of all characters which can be written $\chi' \circ \chi$.

  Since $\chi(\mathbb{Z})$ is isomorphic to $\mathbb{Z}_d$, all characters on $\chi(\mathbb{Z})$ can be written $x \mapsto x^j$ for $j \in \mathbb{Z}_d$ hence $S = \{\chi^j ; j \in \mathbb{Z}_d\}$.

- Clearly

  $$\text{SEI} (\chi(X)) = \sum_{j \in \mathbb{Z}_d \setminus \{0\}} \text{LP} (\chi^j(X))$$

  Idea: if this subsum is significant $\text{SEI}(X)$ then we can focus on $\chi(X)$ instead of $X$ and work with a smaller support.
Using Multiple Characteristics

**Hypothesis** $H_0$: variable $X$ follows uniform distribution $U$

**Hypothesis** $H_1$: variable $X$ follows distribution $P$

\[
1 - \text{BestAdv}_q(U, P) \sim \exp \left( - \frac{q}{8} \sum_{\chi \in \mathcal{Z} \setminus \{1\}} \text{LP}(\chi(X)) \right)
\]

- let $S$ be a set of characters of small order and different from 1
- let $\chi_1, \ldots, \chi_m$ be a basis of span($S$)
- let $h(X) = (\chi_1(X), \ldots, \chi_m(X))$ be the projection of $X$
- we have

\[
\text{SEI}(h(X)) = \sum_{\chi \in \text{span}(S) \setminus \{1\}} \text{LP}(\chi(X)) \geq \sum_{\chi \in S} \text{LP}(\chi(X))
\]

- if this subsum is significant in $\text{SEI}(X)$, we can focus on $h(X)$
New Metrics

Definition

Given a random variable $X$ over an Abelian group $\mathbb{Z}$ of order $n$ and an integer $d$

$$\text{LP}_{\text{max}}(X) = \max_{\chi \in \hat{G}, \chi \neq 1} \text{LP}(\chi(X))$$

$$\text{LP}_{\text{max}}^d(X) = \max_{\chi \in \hat{G}, \chi \neq 1, \chi^d = 1} \text{LP}(\chi(X))$$

note that $\text{LP}_{\text{max}}(X) = \text{LP}_{\text{max}}^n(X)$
Bound on SEI

Theorem

Let $X$ be a random variable over an Abelian group $\mathbb{Z}$ of order $n$.

- we have
  \[
  \text{SEI}(X) \leq (n - 1) \text{LP}_{\text{max}}(X)
  \]
- Let $h$ be a group homomorphism from $\mathbb{Z}$ to a group $G$ of order $d$. We have
  \[
  \text{SEI}(h(X)) \leq (d - 1) \text{LP}_{\text{max}}^d(X)
  \]

Proof.

- first case is a particular case of the second one for $d = n$ and $h$ set to the identity function
- for any $\chi \in \hat{G}$ we have that $\chi \circ h$ is a character over $\mathbb{Z}$ such that $(\chi \circ h)^d = 1$
Tightness of the Bound

Let $X$ be such that

$$\Pr[X = x] = \frac{1 - \varepsilon}{n} + \varepsilon \times 1_{x=0}$$

let $d$ be an integer and consider characters $\chi$ such that $\chi^d = 1$

- if $\chi \neq 1$ and $\chi$ is of order $d$ then

$$E(\chi(X)) = \frac{1}{n} \sum_x \frac{1 - \varepsilon}{n} \chi(x) + \frac{\varepsilon}{n}$$

$$= \frac{\varepsilon}{n}$$

so $LP(\chi(X)) = \frac{\varepsilon^2}{n^2}$

- we deduce $SEI(X) = (n - 1)LP_{\text{max}}(X)$
1. From Statistical Distance to Chernoff Information

2. Applications

3. Further Extensions
   - Spectral Analysis
   - Seek for the Best Efficient Distinguisher
   - Example of Extended Linear Cryptanalysis
Modulo 2 Analysis — Boolean Case (Simple)

Hypothesis $H_0$: Boolean variable $X$ has expected value $\frac{1}{2}$

Hypothesis $H_1$: Boolean variable $X$ has expected value $\frac{1}{2}(1 - \epsilon)$

\[
\text{SEI}(X) = \text{LP}((-1)^X) = |E((-1)^X)|^2 = \epsilon^2
\]

\[
1 - \text{BestAdv}_q(U, P) \sim \exp \left( -\frac{q}{8} \epsilon^2 \right)
\]

input: $x$

1: $c \leftarrow \sum_{i=1}^{q} X_i$

2: $b \leftarrow 1 \left( \frac{c}{q}^{(\epsilon > 0)} \leq \lambda \right)$

output: $b$

\[
(q - c) \log \frac{1}{2(1+\epsilon)} + c \log \frac{1}{2(1-\epsilon)} \leq 0
\]

\[
\frac{c}{q}^{(\epsilon > 0)} \leq \lambda = \frac{1}{1 - \frac{\log(1-\epsilon)}{\log(1+\epsilon)}} \approx \frac{1}{2} \left( 1 - \frac{\epsilon}{2} \right)
\]
Modulo 2 Analysis — Equivalent Form

**Hypothesis** $H_0$: Boolean variable $X$ has expected value $\frac{1}{2}$

**Hypothesis** $H_1$: Boolean variable $X$ has expected value $\frac{1}{2} (1 - \varepsilon)$

$$1 - \text{BestAdv}_q(U, P) \sim \exp \left( -\frac{q}{8} \text{LP}(X) \right)$$

**input:** $x$

1. $c \leftarrow \sum_{i=1}^{q} (-1)^X$

2. $b \leftarrow 1 \left( \frac{c}{q} \geq (1 - 2\lambda) \right)$

   $1 - 2\lambda \approx \frac{\varepsilon}{2}$

   with previous $\lambda$

**output:** $b$
Modulo 2 Analysis — Vectorial Case

Hypothesis $H_0$: variable $X$ follows uniform distribution $U$

Hypothesis $H_1$: variable $X$ follows distribution $P$

$$1 - \text{BestAdv}_q(U, P) \sim \exp \left( -\frac{q}{8} \sum_{\nu \in \mathbb{Z} \setminus \{0\}} \text{LP} (\nu \cdot X) \right)$$

input: $x$

1: $c \leftarrow \sum_{i=1}^{q} (-1)^{u \cdot X}$

2: $b \leftarrow 1 \left( \frac{c}{q} \geq \left( 1 - 2\lambda \right) \right)$

with previous $\lambda$

output: $b$

- best distinguisher between $u \cdot X$ and unbiased coin
- $1 - B_q \sim \exp \left( -\frac{q}{8} \text{LP}(u \cdot X) \right)$
- assume that $\text{LP}(u)$ is overwhelming in $\text{SEI}(P)$...
Modulo 2 Analysis — Boolean Case 2

**Hypothesis** $H_0$: Boolean variable $X$ has expected value $\frac{1}{2}(1 - \varepsilon)$

**Hypothesis** $H_1$: Boolean variable $X$ has expected value $\frac{1}{2}(1 + \varepsilon)$

\[
C(P_0, P_1) \sim \frac{1}{8\ln 2} \left( \frac{\varepsilon^2}{\frac{1}{2}(1 - \varepsilon)} + \frac{\varepsilon^2}{\frac{1}{2}(1 + \varepsilon)} \right) = \frac{1}{2\ln 2} \times \frac{\varepsilon^2}{1 - \varepsilon^2}
\]

\[
1 - \text{BestAdv}_q \sim \exp \left( -\frac{q}{2} \times \frac{\varepsilon^2}{1 - \varepsilon^2} \right)
\]

**input:** $x$

1. $c \leftarrow \sum_{i=1}^{q} (-1)^x^{(\varepsilon>0)}$

2. $b \leftarrow 1 \begin{cases} c \geq 0 \end{cases}$

**output:** $b$
Modulo 2 Analysis — Boolean (Composite)

Hypothesis $H_0$: Boolean variable $X$ has expected value $\frac{1}{2}$

Hypothesis $H_1$: Boolean variable $X$ has expected value $\frac{1}{2}(1 \pm \varepsilon)$

$$1 - \text{BestAdv}_q \sim \exp\left(-\frac{q}{8}\varepsilon^2\right)$$

input: $x$

1. $c \leftarrow \sum_{i=1}^{q} (-1)^X$

2. $b \leftarrow 1 \left(\frac{c}{q} \notin [1 - 2\lambda_- , 1 - 2\lambda_+ \right)$

with $\lambda_\pm = \frac{1}{1 - \log(1 \pm |\varepsilon|)}$

output: $b$

- $1 - 2\lambda_\pm \approx \pm \frac{|\varepsilon|}{2}$
- the test is roughly $\left|\frac{c}{q}\right| \geq \frac{|\varepsilon|}{2}$
Modulo 4 Analysis — Simple

**Hypothesis** $H_0$: variable $X \in \mathbb{Z}_4$ has uniform distribution $U$

**Hypothesis** $H_1$: variable $X \in \mathbb{Z}_4$ has distribution $P$

$$SEI(X) = \sum_{u=1}^{3} |E(i^{u}X)|^2$$

$$1 - \text{BestAdv}_q(U, P) \sim \exp\left(-\frac{q}{8} \text{SEI}(X)\right)$$

**input:** $x$

1: $f_x \leftarrow \frac{1}{q} \sum_{j=1}^{q} 1(X_j = x)$

2: $b \leftarrow 1\left(\sum_x f_x \log \frac{1}{4P(x)} \leq 0\right)$

**output:** $b$

- either a vector of 4 counters
- or a floating point accumulator
Hypothesis $H_0$: variable $X \in \mathbb{Z}_4$ has uniform distribution $U$

Hypothesis $H_1$: variable $X \in \mathbb{Z}_4$ has distribution $P_u$ for $u$ unknown

defined by $P_u(x) = \frac{1-\varepsilon}{4} + \varepsilon \times 1(x = u)$

\[ \forall v \neq 0 \quad \text{LP}(vX) = \left| E(i^{vX}) \right|^2 = \left| \varepsilon i^{vX} \right|^2 = \varepsilon^2 \]

\[ \text{SEI}(X) = 3\varepsilon^2 \]

\[ 1 - \text{BestAdv}_q \sim \exp \left( -\frac{3q}{8}\varepsilon^2 \right) \]

input: $x$

1: $f_x \leftarrow \frac{1}{q} \sum_{j=1}^{q} 1(X_j = x)$

2: $b \leftarrow 1 \left( \min_u \sum_x f_x \log \frac{P_u(x)}{P(x)} \leq 0 \right)$

output: $b$

\[ b = 1 \text{ iff } \max_x f_x \text{ is higher than } \lambda \]

\[ \lambda = \frac{\log(1 - \varepsilon)}{\log(1 - \varepsilon) - \log(1 + 3\varepsilon)} \]
Modulo $d$ Analysis: Generalization — i

**Hypothesis $H_0$:** variable $X \in \mathbb{Z}_d$ has uniform distribution $U$

**Hypothesis $H_1$:** variable $X \in \mathbb{Z}_d$ has distribution $P_u$ for $u$ unknown defined by $P_u(x) = \frac{1-\varepsilon}{d} + \varepsilon \times 1(x = u)$

$$\forall v \neq 0 \quad \text{LP}(vX) = \varepsilon^2$$

$$\text{SEI}(X) = (d-1)\varepsilon^2$$

$$1 - \text{BestAdv}_q \sim \exp \left( -\frac{q(d-1)}{8} \varepsilon^2 \right)$$

- threshold $\lambda$:

```
input: x
1: $f_x \leftarrow \frac{1}{q} \sum_{j=1}^{q} 1(X_j = x)$
2: $b \leftarrow 1(\max_x f_x \geq \lambda)$
output: b
```

$$\frac{\log(1 - \varepsilon)}{\log(1 - \varepsilon) - \log(1 + (d-1)\varepsilon)}$$

so $\lambda \approx \frac{1}{d} + \frac{1-\frac{1}{d}}{2} \varepsilon$

- the optimal test is *not* of form $|\sum_x f_x i^x| \geq \tau$
Modulo 4 Analysis — An Odd Example — i

Consider $X \in \mathbb{Z}_4^{r+1}$

**Hypothesis $H_0$:** variable $X$ has uniform distribution $U$

**Hypothesis $H_1$:** variable $X$ has distribution $P$ induced by

1. pick $x_1, \ldots, x_r$ uniformly at random in $\{0, 1, 2, 3\}$
2. pick $b \in \{0, 1\}$ at random
3. take $x_{r+1} = b + x_1 + \cdots + x_r$

\[
\text{msb}(x_{r+1}) = \bigoplus_{i=1}^{r} \text{msb}(x_i) \oplus \text{msb} \left( b + \sum_{i=1}^{r} \text{lsb}(x_i) \right)
\]

4. let $X = (x_1, \ldots, x_{r+1})$

**Lemma**

\[
\text{msb} \left( \sum_{i=1}^{r} \text{mod } 4 \right) = \bigoplus_{i=1}^{r} \bigoplus_{j=1}^{r} b_i b_j
\]
we can prove that if we represent $X$ as $\tilde{X} \in \mathbb{Z}_2^{2(r+1)}$ with msb’s and lsb’s, we have $\max_{\tilde{u}} \text{LP}(\tilde{u} \cdot \tilde{X}) = 2^{-(r+1)}$ for all $\tilde{u} \neq 0$

$$\text{SEI}(X) = \text{SEI}(\tilde{X}) = \sum_{\tilde{u} \in \mathbb{Z}_2^{2(r+1)}} \text{LP}(\tilde{u} \cdot \tilde{X}) = 1$$

$(2^{r+1} \text{ masks with LP} = 2^{-(r+1)})$

- in $\mathbb{Z}_4^{r+1}$, for $u = (1, 1, \ldots, 1, -1)$ we have $\text{LP}(u \cdot X) = \frac{1}{2}$

$$\text{SEI}(X) = \sum_{u \in \mathbb{Z}_4^{r+1}} \text{LP}(u \cdot X) = 1$$

$(2 \text{ masks with LP} = \frac{1}{2})$  

$\rightarrow$ big gap between modulo 2 and modulo 4 linear cryptanalysis
Modulo $d$ Analysis: Generalization — ii

**Hypothesis** $H_0$: variable $X \in \mathbb{Z}_d$ has uniform distribution $U$

**Hypothesis** $H_1$: variable $X \in \mathbb{Z}_d$ has unknown distribution $P$ with known $\text{SEI}(X)$

$$1 - \text{BestAdv}_q \sim \exp\left(-\frac{q}{8}\text{SEI}(X)\right)$$

**Input:** $x$

1: $f_x \leftarrow \frac{1}{q} \sum_{j=1}^{q} 1(X_j = x)$

2: $\chi^2 \leftarrow d \sum_x \left(f_x - \frac{1}{d}\right)^2$

3: $b \leftarrow 1(\chi^2 \geq \frac{1}{4}\text{SEI}(X))$

**Output:** $b$

The optimal test is neither of form $|\sum_x f_x i^x| \geq \tau$ nor of form $\max_x f_x \geq \lambda$