Today we begin with an example of a probabilistic encryption PKC (public key cryptosystem) which is indistinguishably secure under the assumption that trapdoor functions exist.

Recall that a collection of trapdoor permutations is a set $F = \{f_i : D_i \rightarrow D_i\}_{i \in I}$ such that:

1. $S_1(1^k)$ samples $(i, t_i)$ where $i \in I$, $|i| = k$ and $|t_i| < p(k)$ for some polynomial $p$.
2. $S_2(i)$ samples $x \in D_i$.
3. $\exists$ PTM $A_1$ such that $A_1(i, x) = f_i(x)$.
4. $\Pr[A(i, f_i(x)) \in f_i^{-1}(f_i(x))] < \frac{1}{Q(k)} \forall$ PTM $A, \forall Q, \forall k > k_0$.
5. $\exists$ PTM $A_2$ such that $A_2(i, t_i, f_i(x)) = x, \forall x \in D_i, i \in I$.

Further, let $B_i(x)$ be hard core for $f_i(x)$.

We define a probabilistic encryption PKC based on $F$ to be $(G, E, D)$ where:

1. $G(1^k)$ chooses $(i, t_i)$ by running $S_1(1^k)$ (Public key is $i$, private key is $t_i$).
2. Let $m = m_1 \ldots m_k$ where $m_j \in \{0, 1\}$ be the message.
   
   $E(i, m)$ encrypts $m$ as follows:
   
   - Choose $x_j \in D_i$ such that $B_i(x_j) = m_j$ for $j = 1, \ldots, k$.
   - Output $c = f_i(x_1) \ldots f_i(x_k)$.
3. Let $c = y_1 \ldots y_k$ where $y_i \in D_i$ be the cyphertext.

   $D(i, c)$ decrypts $c$ as follows:
   
   - Compute $m_j = B_i(f_i^{-1}(y_j))$ for $j = 1, \ldots, k$.
   - Output $m = m_1 \ldots m_k$.

**Claim 1** If $F$ is a collection of trapdoor permutations then the probabilistic encryption PKC $(G, E, D)$ is indistinguishably secure.

**Proof:**
Suppose that $(G, E, D)$ is not indistinguishably secure. Then there is a polynomial $Q$, a PTM $A$ and a message space algorithm $M$ such that for infinitely many $k$, $\exists m_0, m_1 \in M(1^k)$ with,

$$\Pr[A(1^k, i, m_0, m_1, c) = j | m_j \in \{m_0, m_1\}, c \in E(i, m_j))] > \frac{1}{2} + \frac{1}{Q(k)}$$

where the probability is taken over the coin tosses of $A$, $(i, t_i) \in G(1^k)$, the coin tosses of $E$ and $m_j \in \{m_0, m_1\}$.

---

1These notes were scribed by Frank D’Ippolito.
In other words, $A$ says 0 more often when $c$ is an encryption of $m_0$ and says 1 more often when $c$ is an encryption of $m_1$.

Define distributions $D_j = E(i, s_j)$ for $j = 0, 1, \ldots, k$ where $k = |m_0| = |m_1|$ and such that $s_0 = m_0$, $s_k = m_1$ and $s_j$ differs from $s_{j+1}$ in precisely 1 bit.

Let $P_j = \Pr[A(1^k, i, m_0, m_1, c) = 1|c \in D_j = E(i, s_j)]$.

Then $\frac{1}{2} + \frac{1}{Q(k)} < \Pr[A$ chooses $j$ correctly] = $(1 - P_0)(\frac{1}{2}) + P_k(\frac{1}{2})$.

Hence, $P_k - P_0 > \frac{2}{Q(k)}$ and since $\sum_{j=0}^{k-1}(P_{j+1} - P_j) = P_k - P_0$, $\exists j$ such that $P_{j+1} - P_j > \frac{2}{Q(k)}$.

Now, consider the following algorithm $B$ which takes input $i, f_i(y)$ and outputs 0 or 1. Assume that $s_j$ and $s_{j+1}$ differ in the $l^\text{th}$ bit; that is, $s_{j,t} \neq s_{j+1,t}$ or, equivalently, $s_{j+1,l} = \bar{s}_{j,l}$.

$B$ runs as follows on input $i, f_i(y)$:
1. Choose $y_1, \ldots, y_k$ such that $B_i(y_r) = s_{j,r}$ for $r = 1, \ldots, k$.
2. Let $c = f_i(y_1), \ldots, f_i(y), \ldots, f_i(y_k)$ where $f_i(y)$ has replaced $f_i(y_1)$ in the $l^\text{th}$ block.
3. If $A(1^k, i, m_0, m_1, c) = 0$ then output $s_{j,l}$.
   If $A(1^k, i, m_0, m_1, c) = 0$ then output $s_{j+1,l} = \bar{s}_{j,l}$.

Note that $c \in E(i, s_j)$ if $B_i(y) = s_{j,l}$ and $c \in E(i, s_{j+1})$ if $B_i(y) = s_{j+1,l}$.

Thus, in step 3 of algorithm $B$, outputting $s_{j,l}$ corresponds to $A$ predicting that $c$ is an encoding of $s_j$; in other words, $c$ is an encoding of the string nearest to $m_0$.

**Claim 2** $\Pr[B(i, f_i(y)) = B_i(y)] > \frac{1}{2} + \frac{1}{Q(k)k}$

**Proof:**

$$\Pr[B(i, f_i(y)) = B_i(y)] = \Pr[A(1^k, i, m_0, m_1, c) = 0|c \in E(i, s_j)] \Pr[c \in E(i, s_j)]$$

$$+ \Pr[A(1^k, i, m_0, m_1, c) = 1|c \in E(i, s_{j+1})] \Pr[c \in E(i, s_{j+1})]$$

$$\geq (1 - P_j)(\frac{1}{2}) + P_k(\frac{1}{2})$$

$$= \frac{1}{2} + \frac{1}{2}(P_{j+1} - P_j)$$

$$> \frac{1}{2} + \frac{1}{Q(k)k}$$

Thus, $B$ will predict $B_i(y)$ given $i$, $f_i(y)$ with probability better than $\frac{1}{2} + \frac{1}{Q(k)k}$. This contradicts the assumption that $B_i(y)$ is hard core for $f_i(y)$.

Hence, the probabilistic encryption PKC $(G, E, D)$ is indistinguishably secure.

In fact, the probabilistic encryption PKC $(G, E, D)$ is also semantically secure. This follows from the fact that semantic and indistinguishable security are equivalent as will be shown in the following theorem.

**Theorem 1** A PKC is semantically secure if and only if it is indistinguishably secure.

**Proof:**

$\implies$ Let $(G, E, D)$ be a semantically secure PKC.
Suppose that \((G, E, D)\) is not indistinguishably secure. Then there is a polynomial \(Q\), a PTM \(A\) and a message space algorithm \(M\) such that for infinitely many \(k\), \(\exists m_0, m_1 \in M(1^k)\) with,

\[
\Pr[A(1^k, P, m_0, m_1, \alpha) = j | m_j \in \{ m_0, m_1 \}, \alpha \in E(P,m_j)] > \frac{1}{2} + \frac{1}{Q(k)}
\]

where the probability is taken over the coin tosses of \(A\), \((P, S) \in G(1^k)\), the coin tosses of \(E\) and \(m_j \in \{ m_0, m_1 \}\).

Construct a new message space \(\hat{M}\) such that for those \(k\) mentioned above,

\[
\Pr[\hat{M}(1^k) = m_0] = \frac{1}{2} = \Pr[\hat{M}(1^k) = m_1].
\]

Consider the information function \(h(m) = m\) and the algorithm \(\hat{A}\) which runs as follows on input \(\alpha \in E(P,m)\):

1. Run \(A(1^k, P, m_0, m_1, \alpha) = j\).
2. Output \(m_j\).

Then, \(\Pr[\hat{A}(1^k, P, \alpha) = h(m) | \alpha \in E(P,m)] = \Pr[A(1^k, P, m_0, m_1, \alpha) = j | m_j \in \{ m_0, m_1 \}, \alpha \in E(P,m_j)] > \frac{1}{2} + \frac{1}{Q(k)}\).

On the other hand, \(\forall\) PTM \(\hat{B}\) (which do not have \(\alpha \in E(P,m)\) as input) we have

\[
\Pr[\hat{B}(1^k) = h(m)] = \Pr[\hat{B}(1^k) = h(m_0) | \hat{M}(1^k) = m_0] \Pr[\hat{M}(1^k) = m_0] + \Pr[\hat{B}(1^k) = h(m_1) | \hat{M}(1^k) = m_1] \Pr[\hat{M}(1^k) = m_1] = \frac{1}{2} \Pr[\hat{B}(1^k) = h(m_j) | \hat{M}(1^k) = m_j] \leq \frac{1}{2}.
\]

Hence, \(\forall\) PTM \(\hat{B}\), \(\Pr[\hat{A}(1^k, P, \alpha) = h(m) | \alpha \in E(P,m)] > \Pr[\hat{B}(1^k) = h(m)] + \frac{1}{Q(k)}\) so that the PKC \((G, E, D)\) fails to satisfy the requirements of semantic security. This is a contradiction.

\(\Leftarrow\) Let \((G, E, D)\) be a indistinguishably secure PKC.

Suppose that \((G, E, D)\) is not semantically secure. Then there is a message space algorithm \(M\), an information function \(h : M \rightarrow V\) and a PTM \(A\) such that \(\forall\) PTM \(B\), \(\exists\) a polynomial \(Q\) and infinitely many \(k\) with,

\[
\Pr[A(1^k, P, \alpha) = h(m)] > \Pr[B(1^k, P) = h(m)] + \frac{1}{Q(k)} \quad (*)
\]

where the first probability is taken over the coin tosses of \(A\), \((P, S) \in G(1^k)\), the coin tosses of \(E\), \(\alpha \in E(P,m)\) and \(m \in M(1^k)\) and the second probability is taken over the coin tosses of \(B\), \((P, S) \in G(1^k)\) and \(m \in M(1^k)\).

However, what we prove here is a slightly different version of the required result; namely, we will show that for all polynomials \(Q\) and PTM’s \(A\) there exists \(k_0\) such that for all \(k \geq k_0\),
Pr[A(1^k, P, \alpha) = h(m)] \leq p_h + \epsilon_k

where \epsilon_k = \frac{1}{Q(k)} and \ p_h = \max_{v \in V} \Pr[h(m) = v] with the probability taken over m \in M(1^k). Note the difference with the definition of semantic security. Essentially, we have substituted \ p_h = \max_{v \in V} \Pr[h(m) = v] for \max_B \Pr[B(1^k, P) = h(m)]. This is exactly the same definition for those h for which it is easy to compute a value v for which \Pr[h(m) = v] = p_h. In such a case, the algorithm B which simply outputs v achieves probability \ p_h. However, the general definition of semantic security is for any h.

Hence, we will assume for contradiction that there is a polynomial Q and a PTM A such that for infinitely many k,

\Pr[A(1^k, P, \alpha) = h(m)] > p_h + \epsilon_k \quad (*)

Let \ r_{m,v} = \Pr[A(1^k, P, \alpha) = v | \alpha \in E(P, m)] = \Pr[A \text{ outputs } v, \text{ given an encryption of } m] \quad \text{and let } \ p_m = \Pr[M(1^k) = m] = \Pr[\text{message } m \text{ is generated}].

Then, (*) can be written as \sum_{m \in M(1^k)} r_{m,h(m)}p_m > p_h + \epsilon_k.

Now, fix a message \mu \in M(1^k).

Let \bar{M}(1^k) = \{m \in M(1^k) : |r_{m,v} - r_{\mu,v}| > \frac{\epsilon_h^2}{10} \text{ for some } v \in V\}.

**Lemma 1** There is an algorithm running in time polynomial in k and \delta^{-1} which on input 1^k, P and m \in \bar{M}(1^k) finds with probability 1 - \delta, v \in V such that |r_{m,v} - r_{\mu,v}| > \frac{\epsilon_h^2}{20}.

**Proof:**

The following is an algorithm which runs as required:

1. Construct a random sample of encodings of m; say, \{\beta_1, \ldots, \beta_s | \beta_j \in E(P, m)\} and let \ I_v(\beta) = \begin{cases} 1 & \text{if } A(1^k, P, \beta) = v \\ 0 & \text{if } A(1^k, P, \beta) \neq v \end{cases}.

   \forall v \in V, \text{ let } q_v = \frac{\sum_{j=1}^{s} I_v(\beta_j)}{s} \text{ be the proportion of times that } A \text{ outputs } v \text{ from the sample. Note that there are at most } s \text{ values of } v \in V \text{ for which } q_v \neq 0.

2. Construct a random sample of encodings of \mu; say, \{\gamma_1, \ldots, \gamma_s | \gamma_j \in E(P, m)\} and let \ I_v(\gamma) = \begin{cases} 1 & \text{if } A(1^k, P, \gamma) = v \\ 0 & \text{if } A(1^k, P, \gamma) \neq v \end{cases}.

   \forall v \in V, \text{ let } \hat{q}_v = \frac{\sum_{j=1}^{s} I_v(\gamma_j)}{s}.

3. If among those v for which either \ I_v(\beta_j) \neq 0 or \ I_v(\gamma_j) \neq 0 \text{ for some } j, \exists \bar{v} \text{ such that } |q_v - \hat{q}_v| > \frac{3\epsilon^2}{40} \text{ then output } \bar{v}.  

4
To see that this output is correct with probability $1 - \delta$ for an appropriate choice of sample size $s$, set $s = \lceil \frac{1}{4(\frac{1}{2})^2} \rceil$ and consider those $v$ for which $|r_{m,v} - r_{\mu,v}| > \frac{\epsilon_k^2}{10}$. (Note that such a $v$ must exist if $m \in \bar{M}(1^k).$)

The weak law of large numbers guarantees that

$$\Pr[|q_v - r_{m,v}| \geq \frac{\epsilon_k^2}{80}] \leq \left(\frac{1}{4}\right) \frac{1}{\left(\frac{\epsilon_k^2}{80}\right)^2} s \leq \left(\frac{1}{4}\right) \frac{(\frac{\delta}{2}) \left(\frac{\epsilon_k^2}{80}\right)^2}{\left(\frac{\epsilon_k^2}{80}\right)^2} = \frac{\delta}{2}$$

so that $\Pr[|q_v - r_{m,v}| < \frac{\epsilon_k^2}{80}] > 1 - \frac{\delta}{2}$ and similarly $\Pr[|\hat{q}_v - r_{\mu,v}| < \frac{\epsilon_k^2}{80}] > 1 - \frac{\delta}{2}$.

Hence, $\Pr[|q_v - \hat{q}_v| > \frac{3\epsilon_k^2}{40}] > \Pr[|q_v - r_{m,v}| < \frac{\epsilon_k^2}{80}$ and $|\hat{q}_v - r_{\mu,v}| < \frac{\epsilon_k^2}{80}]$ (because $|r_{m,v} - r_{\mu,v}| > \frac{\epsilon_k^2}{10}$)

$$= \Pr[|q_v - r_{m,v}| < \frac{\epsilon_k^2}{80}] \Pr[|\hat{q}_v - r_{\mu,v}| < \frac{\epsilon_k^2}{80}]$$

$$> (1 - \frac{\delta}{2})^2 > 1 - \delta$$

This proves Lemma 1.

Lemma 2 $\sum_{m \in \bar{M}} p_m > \frac{\epsilon_k}{10}$.

Proof:

Let $V' = \{v \in V | r_{\mu,v} > \frac{\epsilon_k}{6}\}$ and let $V'' = \{v \in V | r_{\mu,v} \leq \frac{\epsilon_k}{6}\}$.

Let $M' = \{m \in \bar{M} - \bar{M} | r_{\mu,h(m)} > \frac{\epsilon_k}{6}\}$ and let $M'' = \{m \in \bar{M} - \bar{M} | r_{\mu,h(m)} \leq \frac{\epsilon_k}{6}\}$.

Equivalently, $M' = \{\text{messages } m \notin \bar{M} \text{ such that } h(m) \in V'\}$ and $M'' = \{\text{messages } m \notin \bar{M} \text{ such that } h(m) \in V''\}$.

Since $\sum_{v \in V} r_{\mu,v} = 1$, we have that $|V'| < \frac{6}{\epsilon_k}$.

Now, $p_h + \epsilon_k < \sum_{m \in \bar{M}(1^k)} r_{m,h(m)}p_m = \sum_{m \in \bar{M}} r_{m,h(m)}p_m + \sum_{m \in \bar{M} - \bar{M}} r_{m,h(m)}p_m$.

By the definition of $\bar{M}$, $r_{m,h(m)} < r_{\mu,h(m)} + \frac{\epsilon_k^2}{10}$ for $m \in \bar{M} - \bar{M}$.

Thus, $p_h + \epsilon_k < \sum_{m \in \bar{M}} r_{m,h(m)}p_m + \sum_{m \in \bar{M} - \bar{M}} (r_{\mu,h(m)} + \frac{\epsilon_k^2}{10})p_m$

$$\leq \sum_{m \in \bar{M}} r_{m,h(m)}p_m + \sum_{m \in M'} (r_{\mu,h(m)} + \frac{\epsilon_k^2}{10})p_m + \sum_{m \in M''} (\frac{\epsilon_k}{6} + \frac{\epsilon_k^2}{10})p_m$$

$$\leq \sum_{m \in \bar{M}} r_{m,h(m)}p_m + \sum_{v \in V'} \sum_{m \in h^{-1}(v)} (r_{\mu,v} + \frac{\epsilon_k^2}{10})p_m + (\frac{\epsilon_k}{6} + \frac{\epsilon_k^2}{10})(1)$$

5
But \( \sum_{v \in V'} \sum_{m \in h^{-1}(v)} (r_{\mu,v} + \frac{\epsilon_k^2}{10})p_m = \sum_{v \in V'} \sum_{m \in h^{-1}(v)} p_m + \sum_{v \in V'} \sum_{m \in h^{-1}(v)} \frac{\epsilon_k^2}{10}p_m \)
\[ \leq (\sum_{v \in V'} r_{\mu,v})(\max_{v \in V'} \sum_{m \in h^{-1}(v)} p_m) + (\frac{\epsilon_k^2}{10} \sum_{v \in V'} (\max_{v \in V'} \sum_{m \in h^{-1}(v)} p_m) \]
\[ \leq (1)(p_h) + (\frac{\epsilon_k^2}{10})(\frac{80}{5})(p_h) \]
\[ \leq p_h + \frac{3\epsilon_k}{5} \]

Hence, \( p_h + \epsilon_k < \sum_{m \in M} r_{m,h(m)}p_m + p_h + \frac{3\epsilon_k}{5} + \frac{\epsilon_k}{6} + \frac{\epsilon_k^2}{10} \)
\[ \leq \sum_{m \in M} 1 \cdot p_m + p_h + \frac{3\epsilon_k}{5} + \frac{\epsilon_k}{6} + \frac{\epsilon_k^2}{10} \quad \text{(without loss of generality } \epsilon_k \leq 1) \]

and thus \( \epsilon_k < \sum_{m \in M} p_m + \frac{13\epsilon_k}{15} \) or \( \sum_{m \in M} p_m > \frac{2\epsilon_k}{15} > \frac{\epsilon_k}{10} \) which proves Lemma 2. \( \square \)

Now, consider the algorithm \( F \) which runs as follows:

1. Fix an element \( \mu \in M \).
2. Randomly generate an element \( m \in M \). Note that \( m \in \tilde{M} \) with probability at least \( \frac{\epsilon_k}{10} \) by Lemma 2.
3. Using the process described in Lemma 1, search for an element \( v \in V \) such that \( |r_{m,v} - r_{\mu,v}| > \frac{\epsilon_k^2}{20} \). Note that if \( m \in \tilde{M} \) then \( F \) succeeds with high probability by Lemma 1.
4. If no \( v \) is found in step 3 then repeat from step 2.

By Lemmas 1 and 2, \( F \) succeeds in expected polynomial time. Let \( \{m_0, m_1\} = \{\mu, m\} \) in such a way that \( r_{m_1,v} > r_{m_0,v} + \frac{\epsilon_k^2}{20} \). Note that \( F \) can determine this with high probability because \( q_v \) and \( \tilde{q}_v \) from Lemma 1 approximate \( r_{m,v} \) and \( r_{\mu,v} \), respectively, to within \( \frac{\epsilon_k^2}{20} \) with high probability.

Finally, consider the algorithm \( \hat{A} \) which runs as follows on inputs \( m_0, m_1 \) and \( \alpha \in E(P,m_j) \) where \( j \in \{0,1\} \):

1. Run \( A(1^k, P, \alpha) = u \).
2. If \( u = v \) then output 1.
3. If \( u \neq v \) then output \( j \in \{0,1\} \) randomly.

Then \( \Pr[\hat{A}(1^k, P, m_0, m_1, \alpha) = j | m_j \in \{m_0, m_1\} \alpha \in E(P,m_j)] = \Pr[j = 1] \Pr[\hat{A}(1^k, P, m_0, m_1, \alpha) = 1 | \alpha \in E(P,m_1)] + \Pr[j = 0] \Pr[\hat{A}(1^k, P, m_0, m_1, \alpha) = 0 | \alpha \in E(P,m_0)] \)
\[ = \frac{1}{2} \left( \sum_{u \neq v} \Pr[\hat{A}(1^k, P, \alpha) = u | \alpha \in E(P,m_1)] \cdot \frac{1}{2} + \Pr[\hat{A}(1^k, P, \alpha) = v | \alpha \in E(P,m_1)] \cdot 1 \right) \]
\[ + \frac{1}{2} \left( \sum_{u \neq v} \Pr[\hat{A}(1^k, P, \alpha) = u | \alpha \in E(P,m_0)] \cdot \frac{1}{2} + \Pr[\hat{A}(1^k, P, \alpha) = v | \alpha \in E(P,m_0)] \cdot 0 \right) \]
6
\[
\frac{1}{2} \left( (1 - r_{m_1,v}) \cdot \frac{1}{2} + r_{m_1,v} \cdot 1 \right) + \frac{1}{2} \left( (1 - r_{m_0,v}) \cdot \frac{1}{2} + r_{m_0,v} \cdot 0 \right)
\]
\[
> \frac{1}{2} + \frac{2^k}{80}
\]

This holds for the infinitely many \( k \) mentioned at the outset and so contradicts the requirements of indistinguishable security. This completes the proof of the theorem. 

The following alternate proof that indistinguishable security implies semantic security is due to Michael Klugerman. Throughout this proof, all notation is as before except that the definition of \( \bar{M}(1^k) \) for fixed \( \mu \in M(1^k) \) is modified as follows:

\[
\bar{M}(1^k) = \{ m \in M(1^k) : |r_{m,v} - r_{\mu,v}| > \frac{1}{Q(k)} \text{ for some } v \in V \text{ and some polynomial } Q \}.
\]

Proof:
Consider a PKC which is indistinguishably secure. Then the algorithm \( F \) described above cannot succeed in expected polynomial time and since the consequence of Lemma 1 is still valid it must be the case that the following property holds:

Property 1

\[
\forall \text{message space algorithms } M, \forall \text{PTM } A \text{ and } \forall \text{polynomials } Q, \exists k_0 \text{ such that for } k \geq k_0
\]
\[
\sum_{m \in \bar{M}} p_m < \frac{1}{Q(k)}
\]

(There is, however, a subtlety involved here. In order for the proof of Lemma 1 to be valid, we must work with a fixed value of \( \epsilon_k = \frac{1}{Q(k)} \). To see that such a value is available, note that \( \bar{M} \) is finite and that for each \( m_i \in \bar{M} \), \( \exists \) a polynomial \( Q_i \) such that \( |r_{m_i,v} - r_{\mu,v}| > \frac{1}{Q_i(k)} \) for some \( v \). Consequently, if \( Q \) is chosen so that \( Q(k) = \max_i Q_i(k) \) then \( |r_{m,v} - r_{\mu,v}| > \frac{1}{Q(k)} \) for all \( m \in \bar{M} \) and the proof of Lemma 1 goes through.)

Next we show that this property guarantees semantic security. Consider a PKC with Property 1 and the following algorithm \( B \):

1. Randomly generate an element \( m \in M(1^k) \).
2. Encrypt \( m \), generating \( \alpha = E(P, m) \).
3. Output \( A(1^k, P, \alpha) \).

Analogous to \( r_{m,v} \), define \( s_{m,h(m)} = \Pr[B(1^k, P) = h(m)] \). To show that the PKC is semantically secure, it suffices to show that for any polynomial \( R \),

\[
\sum_{m \in \bar{M}} p_m r_{m,h(m)} - \sum_{m \in M} p_m s_{m,h(m)} \leq \frac{1}{R(k)}.
\]
Now, \[ \sum_{m \in M} p_m r_{m,h}(m) - \sum_{m \in M} p_m s_{m,h}(m) = \sum_{m \in M} p_m (r_{m,h}(m) - s_{m,h}(m)) \]
\[ = \sum_{m \in M} p_m (r_{m,h}(m) - s_{m,h}(m)) + \sum_{m \notin \bar{M}} p_m (r_{m,h}(m) - s_{m,h}(m)) \]
\[ \leq \sum_{m \in M} p_m + \max_{m \notin \bar{M}} (r_{m,h}(m) - s_{m,h}(m)) \]

By Property 1, \[ \sum_{m \in M} p_m < \frac{1}{2R(k)} \] for any polynomial \( R \). Thus, it remains to show that this is also the case for \[ \max_{m \notin \bar{M}} (r_{m,h}(m) - s_{m,h}(m)) \]. Consider the value \( m_{\max} \notin \bar{M} \) which maximizes this quantity. By the construction of \( B \),
\[ s_{m_{\max},h}(m_{\max}) = \sum_{m \in M} p_m r_{m,h}(m_{\max}) \]
\[ = \sum_{m \in M} p_m r_{m,h}(m_{\max}) + \sum_{m \notin \bar{M}} p_m r_{m,h}(m_{\max}) \]
\[ \geq \sum_{m \notin \bar{M}} p_m \left( r_{m_{\max},h}(m_{\max}) - \frac{1}{4R(k)} \right) \]

This last inequality holds because \( m, m_{\max} \notin \bar{M} \) so that for any polynomial \( R \),
\[ |r_{m,h}(m_{\max}) - r_{m_{\max},h}(m_{\max})| \leq |r_{m,h}(m_{\max}) - r_{\mu,h}(m_{\max})| + |r_{\mu,h}(m_{\max}) - r_{m_{\max},h}(m_{\max})| \]
\[ \leq \frac{1}{8R(k)} + \frac{1}{8R(k)} = \frac{1}{4R(k)} \]
from which \( r_{m,h}(m_{\max}) - r_{m_{\max},h}(m_{\max}) \geq -\frac{1}{4R(k)} \) for any polynomial \( R \).

Hence, \[ s_{m_{\max},h}(m_{\max}) \geq \left( r_{m_{\max},h}(m_{\max}) - \frac{1}{4R(k)} \right) \left( \sum_{m \notin \bar{M}} p_m \right) \]
\[ \geq \left( r_{m_{\max},h}(m_{\max}) - \frac{1}{4R(k)} \right) \left( 1 - \frac{1}{4R(k)} \right) \quad \text{(for any } R \text{ by Property 1)} \]
\[ \geq r_{m_{\max},h}(m_{\max}) - \frac{1}{2R(k)} \]

Hence the desired inequality follows, proving the result.
References