Identity of King and Flajolet & al. Formulae for LRU Miss Rate
Exact Computation

Christian BERTHET
STMicroelectronics, Grenoble, France,

Abstract. This short paper gives a detailed proof of identity between two classic formulas for the computation of the exact Miss Rate of LRU caches. An extension to the identity of two formulas of the expected time of a partial collection in the coupon collector problem is also presented.

Keywords: LRU (Least-Recently Used) Cache, Miss Rate Exact computation, Coupon Collector Problem.

Address all correspondence to: BERTHET Christian; E-mail: Christian.berthet@st.com

Introduction

There exist two formulae for the exact computation of LRU Miss Rate. As early as 1971 W.F. King derived a formula to compute the fault probability (‘miss rate’) of LRU demand paging systems [King71]. And, in 1992, Flajolet et al. gave an integral formula of that probability [Flajolet92]. In their paper, Flajolet et al. mention that “King gave another form of that probability”, implying that they used a different method for counting the same thing. However, to our knowledge, a formal equivalence between these two formulae has never been established yet. In this paper we show a direct equivalence between the two by algebraic means.

In both cases, hypotheses on the LRU system are the following: system is assumed to obey an Independent-Reference Model (IRM), i.e. cache accesses are independent of past history. Addresses of cache lines (or ‘pages’) are characterized by a ‘popularity’ distribution, i.e. a general probability law (not necessarily uniform). Finally, for readers familiar with HW design, the cache is fully associative (in other words, dealing with limited associativity such as in HW caches require the need of resorting to other models).

In this short paper, we show the equivalence of the two afore-mentioned LRU Miss Rate formulas. Proof is based on a simple rewriting mechanism that makes use of a combinatorial identity on the permutations of a subset of probabilities proven in Lemma 1.

In a final section, we give an extension to the equivalence of two other, and similar, formulas for the exact computation of the expected time of a partial collection in the coupon collector problem under a general probability law.
King Formula for LRU MR

Notation

We assume a probability law with general (non-uniform) distribution \( p_i, i \in \{1,...,m\}, \sum_{i=1}^{m} p_i = 1 \).
Probability law is often called ‘popularity’ of the addresses \{1,...,m\}.

Original Formula

King formula for the Miss Rate of an LRU cache of capacity \( j, 1<j<m \) is [King71] as follows

\[
\text{King}[j] = \sum_{j\text{-tuple } \{i_1, i_2,...,i_j\}} \frac{p_{i_1}p_{i_2}...p_{i_{j+1}}(1-p_{i_1}p_{i_2}...p_{i_{j+1}})}{(1-p_{i_1})(1-p_{i_1}-p_{i_2})...(1-p_{i_1}-p_{i_2}...-p_{i_{j+1}})}
\]

Miss rate is a summation over all the \( j \)-tuples of size \( j \), i.e. all the permutations of all the subsets of size \( j \) of the set \{1,..,m\}. This Formula is also given in Fagin papers [Fagin77, Fagin78].

Rewriting of King Formula

Let J be a subset of the set \{1,...,m\}, with \( |J|<m \), we introduce the notation:

\[
I_j = \sum_{\text{permutation } \{i_1, i_2,...,i_j\}\text{of }J} \frac{p_{i_1}p_{i_2}...p_{i_j}}{(1-p_{i_1})(1-p_{i_1}-p_{i_2})...(1-p_{i_1}-p_{i_2}...-p_{i_j})}
\]

Noticing that \( \sum_{i \in J} p_i \) is independent of the permutation of the subset J, King initial formula is rewritten: \( \text{King}[j] = \sum_{|J|=j} \left( 1 - \sum_{i \notin J} p_i \right)^2 I_j = \sum_{|J|=j} \left( \sum_{i \notin J} p_i \right)^2 I_j \).

Flajolet Integral Formula

Original formula

An integral formula of the LRU Miss rate of a cache of size \( j \) is given by Flajolet et al. ([Flojolet92] Theorem 5.1 pg. 219):

\[
1 - \text{MR}[j] = \int_0^\infty e^{-t} \prod_{i=1}^{m} \left( 1 + u(e^{p_i} - 1) \right) \left( \sum_{i=1}^{m} \frac{p_i^2}{1 + u(e^{p_i} - 1)} \right) e^{-t} dt
\]
where \([u^j]f(u)\) denotes the coefficient of \(u^j\) in the polynomial \(f(u)\). It can be re-written using successively variable change \(x = e^{-t}\), distribution of the sum, computation of \(u^j\) coefficient and distribution of the integral:

\[
1 - MR[j] = \sum_{q=0}^{j-1} \sum_{i=1}^{m} \left[ u^q \right] \left( \prod_{i=1}^{m} \frac{p_i^2}{1 + u(x^{-pi} - 1)} \right) \left( \sum_{i=1}^{m} u^q \prod_{k \neq i} (1 + u(x^{-ki} - 1)) \right) dx = \sum_{q=0}^{j-1} \sum_{i=1}^{m} \left[ u^q \right] p_i^2 \left( \prod_{k \neq i} (1 + u(x^{-ki} - 1)) \right) dx
\]

\[
= \sum_{q=0}^{j-1} \left( \sum_{i=1}^{m} p_i^2 \sum_{q \in K \subseteq Q, k \in K} \prod_{k \neq i} (x^{-ki} - 1) \right) dx = \sum_{q=0}^{j-1} \sum_{i=1}^{m} p_i^2 \sum_{q \in K \subseteq Q, k \in K} \prod_{k \neq i} (x^{-ki} - 1) \right) dx
\]

We now define the following lemma which is proven in Appendix 1. It allows us to relate the two MR formulae using the quantity \(I_J\) defined above.

**Lemma 1**

Let \(J\) be a subset \(J\) of size \(j\) of the set \(\{1, \ldots, m\}\), \(1 \leq j < m\). It holds that:

\[
I_J = \int_{0}^{1} \left( \prod_{i \in J} (x^{-pi} - 1) \right) dx
\]

**Rewriting of Flajolet et al. formula**

Using Lemma 1, Flajolet et al. formula is:

\[
1 - MR[j] = \sum_{q=0}^{j-1} \sum_{i=1}^{m} p_i^2 \sum_{q \in K \subseteq Q, k \in K} \prod_{k \neq i} I_k = \sum_{i=1}^{m} p_i^2 \sum_{0 < |K| < j} I_k = \sum_{0 < |K| < j} I_k \left( \sum_{i \in K} p_i^2 \right)
\]

Using the fact that, for \(j=1\), \(1 - MR[1] = \sum_{i=1}^{m} p_i^2\) (obviously a cache of capacity 1 hits only for successive accesses to the same address) \(I_\emptyset\) notation can be extended to null set with: \(I_\emptyset = 1\).

Also since Miss Rate is null for a capacity \(j=m\), it holds that \(\sum_{0 < |K| < m} I_K \left( \sum_{i \in K} p_i^2 \right) = 1\), so MR\([j]\) can also be expressed as follows: \(\sum_{j \leq j \leq m} I_J \left( \sum_{i \in J} p_i^2 \right)\).

At first sight, King formula may seem more tractable than that of Flajolet et al. since the summation is performed on a single value \(j\) and not on a subset \([0,j]\) or \([j,m]\). However this is very arguable since both expressions require the enumeration of all subsets which is intractable.
Equivalence of King and Flajolet formulas

Equivalence of these two formulas reduces to proving that: 
\[ \sum_{|J|=j} \left( \sum_{i \in J} p_i \right)^2 I_j = 1 - \sum_{0 \leq |J| < j} I_j \left( \sum_{i \in J} p_i^2 \right) \]

Uniform Distribution

Let us first reassure ourselves all this makes sense with a simple uniform distribution (\( p_i = 1/m \) for all \( i \)). Using variable change \( X = x^{1/m} \) (i.e., \( mX^{-1}dX = dx \)), it follows that: 
\[ I_j = \int_0^1 (x^{-j/m} - 1)^j dx = \int_0^1 (X^{-1} - 1)^j X^{m-1} dX = m \frac{1}{m!} \frac{j!(m-j-1)!}{(m-1)!} \] 

Relation \( \int_0^1 x^a (1-x)^b dx = \frac{a!b!}{(a+b+1)!} \) is classically obtained using iterated integration by parts.

King Formula collapses to the well-known LRU MR value for a uniform distribution: 
\[ \sum_{|J|=j} \left( \sum_{i \in J} p_i \right)^2 I_j = \binom{m}{j} \left(1 - \frac{j}{m}\right)^2 \left(\frac{m-1}{j}\right) = 1 - \frac{j}{m}. \]

And Flajolet et al. on the other hand: 
\[ 1 - MR[j] = \sum_{0 \leq |J| < j} I_j \left( \sum_{i \in J} p_i^2 \right) = \sum_{q=0}^{j-1} \sum_{|J|=q} I_j \left( \sum_{i \in J} p_i^2 \right) = \sum_{q=0}^{j-1} \binom{m}{q} \left(\frac{m-q}{m}\right)^{m-1} \frac{1}{m} \sum_{q=0}^{j-1} 1 = \frac{j}{m}. \]

Before proving the equivalence in the general case, we first introduce the following lemma which is proven in Appendix 2.

Lemma 2

Let \( J \) be a subset of the set \( \{1, \ldots, m\} \) of size \( j \leq m \), and \( F \) be a numerical function. It holds that: 
\[ \sum_{|J|=j} \sum_{i \in J} F(i, J) = \sum_{|I|=j+1} \sum_{i \in J} F(i, J - \{i\}) \]

Equivalence of King and Flajolet formulas in general case

We give the proof of equivalence for a general distribution: 
\[ \sum_{|J|=j} \left( \sum_{i \in J} p_i \right)^2 I_j = 1 - \sum_{0 \leq |J| < j} I_j \left( \sum_{i \in J} p_i^2 \right) \]

Identity is obvious for \( |J| = 1 \) since: 
\[ \sum_{j=1}^n (1-p_j)^2 \frac{p_j}{1-p_j} = \sum_{i=1}^m p_i - \sum_{i=4}^m p_i^2 = 1 - \sum_{i=1}^m p_i^2. \]
Also note that for j=m-1:  \( \sum_{\{i=1\ldots m-1\}} (p_{ai,j})^2 I_j = \sum_{\{i=1\ldots m\}} I_j (p_{ai,j})^2 \).

For the general case we proceed by induction on the size j:

\[
MR[j+1] = 1 - \sum_{0<j'<j+1} I_{j'} \sum_{i \in J} p_{i}^2 = MR[j] - \sum_{i \in J} I_{j} \sum_{i \in J} p_{i}^2.
\]

From the hypothesis \( MR[j]=\text{King}[j] \) and noting \( q_{j} = \sum_{i \in J} p_{i} \), induction step is:

\[
MR[j+1] = \sum_{\{i=1\ldots j+1\}} \left( \sum_{i \in J} p_{i} \right)^2 I_{j} = \sum_{\{i=1\ldots j+1\}} \left( \sum_{i \in J} p_{i} \right) I_{j} = \sum_{\{i=1\ldots j+1\}} \left( \sum_{i \in J} p_{i} (q_{j} - p_{i}) \right) I_{j}
\]

On the other hand: \( \text{King}[j+1] = \sum_{\{i=1\ldots j+1\}} \left( \sum_{i \in J} p_{i} \right)^2 I_{j} = \sum_{\{i=1\ldots j+1\}} q_{j}^2 I_{j} \)

From the definition of \( I_{j} \) the following relation holds: \( q_{j} I_{j} = \sum_{i \in J} p_{i} I_{j} \). This stems directly from considering the permutations of \( J \) whose last element corresponds to the sum index ‘i’.

Consequently \( \text{King}[j+1] = \sum_{\{i=1\ldots j+1\}} \sum_{i \in J} q_{j} p_{i} I_{j} = \sum_{\{i=1\ldots j+1\}} \sum_{i \in J} (q_{j} - p_{i}) p_{i} I_{j} \). 

Lemma 2 applied to the function \( F(i, J-{i}) = (q_{j} - p_{i}) p_{i} I_{j} \) finally leads to the desired result: \( \text{King}[j+1] = \sum_{\{i=1\ldots j\}} \sum_{i \in J} (q_{j} - p_{i}) p_{i} I_{j} = MR[j+1] \). QED.
Application to CCP expectation of a partial collection

*Flajolet et al. Integral Formula for CCP expectation*

For a general (non-uniform) distribution $p_i, i \in \{1..m\}, \sum_{i=1}^{m} p_i = 1$, Flajolet et al. have given a formula for the expected waiting time to collect $j$ items out of $m$ (‘partial collection’) of the coupon-collector problem:

Expected time is ([Flajolet92] 13a p.216):

$$E(C_j) = \int_0^{+\infty} u^j \left(\prod_{i=1}^{m} (1 + u(e^{p_i} - 1))\right) e^{-u} \, du ,$$

where $j \leq m$ is the size of the collection and $[u^q]f(u)$ denotes the coefficient of $u^q$ in the polynomial $f(u)$. It is obvious that $E(C_0) = 0$ and $E(C_1) = 1$.

Let us notice that formula has a simple recurrence relation:

$$E(C_{j+1}) - E(C_j) = \int_0^{+\infty} u^j \left(\prod_{i=1}^{m} (1 + u(e^{p_i} - 1))\right) e^{-u} \, du , \quad 1 \leq j < m,$$

which can be rewritten with variable change $x = e^t$, computation of $u^j$ coefficient and distribution of the integral: $E(C_{j+1}) - E(C_j) = \sum_{|J|=j} \left(\prod_{i \in J} x^{-p_i} - 1\right) \, dx$.

In the sequel, we denote this expression $\Delta E\{j\}$. Using Lemma 1 and $I_J$ notation, it follows that:

$$\Delta E\{j\} = \sum_{|I|=j} I_J \quad \text{and} \quad E(C_j) = \sum_{0 \leq |J| \leq j} I_J .$$

Let us notice that, for a uniform distribution: $\Delta E\{j\} = m \left(\frac{1}{m}\right) \left(\frac{1}{m-1}\right) \ldots \left(\frac{1}{j}\right) = \frac{m}{m-j}$ giving the well-known

$$E(C_j) = m(H_m - H_{m-j}) \quad \text{where} \, H_n \text{ is the n-th harmonic number.}$$

In their paper, Flajolet et al. also give a symmetric functions expression (or a variant of it, after index change)

$$E(C_j) = \sum_{k=m-j+1}^{m} (-1)^{j-k-m+1} \binom{k-1}{m-j} \sum_{J=\emptyset}^{k} P_J$$

where $P_J$ is the sum of the probabilities of the subset $J$.

*Alternative formula from Ferrante et al.***

An alternative formula for the expectation of a partial collection, noted $E[X_m(k)]$ – equivalent of Flajolet $E(C_k)$ - and based on conditional probabilities is given in Proposition 2 page 7 of paper [Ferrante12]. It is very similar to King Formula for the LRU miss rate calculation since it requires the enumeration of all the permutations of all the subsets. Expectation for a partial
collection of size k out of m is defined as: $E[X_m(k)] = \sum_{i=1}^{k} E[X_s]$ where $E[X_s]$ is (using their notation for the permutations):

$$E[X_s] = \sum_{i_{i_1}\neq i_{i_1} \neq \ldots \neq i_{i_1}}^{m} \frac{P_{i_1} P_{i_2} \cdots P_{i_{i_1}}}{(1 - p_{i_1})(1 - p_{i_1} - p_{i_2}) \ldots (1 - p_{i_1} - p_{i_2} \cdots - p_{i_{i_1}})}$$

Again, there is a very simple recurrence relation $E[X_m(k)] = E[X_m(k-1)] + E[X_k]$ (assuming $E[X_1]=1$).

**Equivalence of Flajolet et al. $E[C_j]$ and Ferrante&al. $E[X_m(j)]$**

For a uniform probability $E[X_s] = \frac{m!}{(m-s+1)!} \frac{(m-s)!}{(m-1)!} = \frac{m}{m-s+1}$, hence

$$E[X_m(k)] = \sum_{i=1}^{k} \frac{m}{m-s+1} = E\{C_k\}.$$

For $j=2$, whatever $m\geq 2$, equivalence is readily obtained by computation of the sums of Flajolet et al. symmetric functions expression:

$$E[C_j] = \sum_{k=m-2}^{m} (-1)^{m-2-k} \left( \begin{array}{c} k-1 \\ m-2-j \end{array} \right) \sum_{|J|=|J|=k} \frac{1}{P_j} = \sum_{|J|=m-1} \frac{1}{P_j} + (-1)^{p-1} \sum_{|J|=m-2} \frac{1}{P_j} = \sum_{i=1}^{m} \frac{1}{1 - p_i} - (m-1)$$

$$= 1 + \sum_{i=1}^{m} \frac{p_i}{1 - p_i} = E[X_m(2)]$$

More generally, identity of Flajolet et al. and Ferrante et al. relations for any $j \leq m$ can be stated simply by noticing that $E[X_s] = \sum_{|J|=s} I_j$ hence $E[X_m(k)] = \sum_{i=1}^{k} \sum_{|J|=s} I_j = E\{C_k\}$.

**Conclusion**

Using algebraic means we have shown the identity of two formulas for the exact computation of LRU miss rate and two other formulas for the expectation of a partial collection in CCP. Both identities make use of a combinatorial relation on the permutations of a subset proved in Appendix 1.

**Acknowledgments**

The author thanks Russell May for his very helpful comments.
References

1. W.F. King III, “Analysis of demand paging algorithms”, Proc. of the I.F.I.P. Congress 1971, North-Holland Publishing Company.

2. Ronald Fagin, “Asymptotic miss ratios over independent references”, Journal of Computer and System Sciences Volume 14, Issue 2, April 1977, Pages 222–250
   http://www.sciencedirect.com/science/article/pii/S0022000077800147

3. Fagin, R., & Price, T. G. (1978). Efficient calculation of expected miss ratios in the independent reference model. SIAM Journal on Computing, 7(3), 288-297.

4. Flajolet, Philippe; Gardy, Danièle; Thimonier, Loïs (1992), "Birthday paradox, coupon collectors, caching algorithms and self-organizing search", Discrete Applied Mathematics 39 (3): 207–229.
   doi:10.1016/0166-218X(92)90177-C, MR 1189469.

5. Ferrante, M., & Frigo, N. (2012). On the expected number of different records in a random sample.
   arXiv preprint arXiv:1209.4592.
Appendix 1: Proof of Lemma 1

Let $J$ be a subset $J$ of size $j$ of the set $\{1, \ldots, m\}$, $1 \leq j < m$. We want to prove that:

$$\int_0^1 \left( \prod_{i \in J} (x^{-p_i} - 1) \right) dx = \sum_{\text{permutation } (i_1, i_2, \ldots, i_j) \in J} \frac{p_{i_1} p_{i_2} \cdots p_{i_j}}{(1 - p_{i_1})(1 - p_{i_1} - p_{i_2}) \cdots (1 - p_{i_j} - p_{i_{j-1}} - \cdots - p_{i_j})}$$

For sake of simplicity of notation, we note $\{1,2,\ldots,j\}$ the subset $J$.

Identity is easily proven for $j=1$ and $j=2$ by expanding and evaluating the integrals.

For $|J|=1$:

$$\int_0^1 (x^{-p_1} - 1) dx = \frac{1}{p_1 - 1}$$

and for $|J|=2$:

$$\int_0^1 (x^{-p_1} - 1)(x^{-p_2} - 1) dx = \frac{1}{1 - p_1 - p_2} - \frac{1}{1 - p_1} - \frac{1}{1 - p_2} + 1 = \frac{p_1 p_2 (2 - p_1 - p_2)}{(1 - p_1)(1 - p_2)(1 - p_1 - p_2)}$$

We prove the general case by induction on $j$ and we note

$$R_j = \int_0^1 \left( \prod_{i \in J} (x^{-p_i} - 1) \right) dx$$

Then: $R_{j+1} = \int_0^1 \left( \prod_{i \in J^\prime \cup \{j+1\}} (x^{-p_i} - 1) \right) dx = \int_0^1 \left( \prod_{i \in J^\prime} (x^{-p_i} - 1) \right) x^{-p_{j+1}} dx - R_j$

We introduce the variable change $z = x^{-p_{j+1}}$ hence $dz = \left( 1 - p_{j+1} \right) x^{-p_{j+1}} dx$ and

$$R_{j+1} = \frac{1}{1 - p_{j+1}} \int_0^1 \left( \prod_{i \in J^\prime} (z^{-1} - 1) \right) dz - \int_0^1 \left( \prod_{i \in J^\prime} (x^{-p_i} - 1) \right) dx$$

At this point we assume the induction hypothesis holds for the two integrals, both of them being indexed by $j$. For the first integral, notice that:

$$\int_0^1 \left( \prod_{i \in J^\prime} \left( z^{-p_{i}} - 1 \right) \right) dz = \sum_{\text{permutation } (i_1, i_2, \ldots, i_j) \in J^\prime} \frac{p_{i_1} \cdots p_{i_j}}{(1 - p_{i_1})(1 - p_{i_1} - p_{i_2}) \cdots (1 - p_{i_j} - p_{i_{j-1}} - \cdots - p_{i_j})}$$

We use the intermediate variables $\{q_{i_1} = 1 - p_{i_1}, q_{i_2} = 1 - p_{i_2} - p_{i_1}, \ldots, q_{i_j} = 1 - p_{i_j} - \cdots - p_{i_1}\}$ to make the proof more readable, hence:

$$R_j = \int_0^1 \left( \prod_{i \in J} (x^{-p_i} - 1) \right) dx = \sum_{\text{permutation } (i_1, i_2, \ldots, i_j) \in J} \frac{p_{i_1} p_{i_2} \cdots p_{i_j}}{q_{i_1} q_{i_2} \cdots q_{i_j}}$$

And:

C.BERTHET  Page 9  7/5/2016
\[ R_{j+1} = \left( \frac{1}{1 - p_{j+1}} \right) \sum_{\text{permutations of } J \text{ of size } j+1} \frac{p_i \cdot p_j \cdot p_{j+1}}{(1 - p_{j+1})(q_i - p_{j+1})(q_j - p_{j+1}) \cdot (q_{j+1} - p_{j+1})} - \sum_{\text{permutations of } J \text{ of size } j} \frac{p_i \cdot p_j \cdot p_{j+1}}{q_i \cdot q_j \cdot q_{j+1}} \]

Now, since in a subset of size j+1, there are (j+1) permutations for each permutation of a subset J of size j, we can enumerate all the possible positions of element j+1:

\[
\sum_{\text{permutations of } J \text{ of size } j} \frac{p_i \cdot p_j \cdot p_{j+1}}{q_i \cdot q_j \cdot q_{j+1}} = \sum_{\text{permutations of } J \text{ of size } j} \left( \frac{p_{j+1} \cdot p_i \cdot p_j}{(1 - p_{j+1})(q_i - p_{j+1})(q_j - p_{j+1}) \cdot (q_{j+1} - p_{j+1})} + \frac{p_i \cdot p_j \cdot p_{j+1}}{q_i \cdot q_j \cdot q_{j+1}} \right) + \sum_{\text{permutations of } J \text{ of size } j} \frac{p_i \cdot p_j \cdot p_{j+1}}{q_i \cdot q_j \cdot q_{j+1}}
\]

Lemma 1 holds if and only if we can prove that \( R_{j+1} \) and \( \sum_{\text{permutations of } J \text{ of size } j} \frac{p_i \cdot p_j \cdot p_{j+1}}{q_i \cdot q_j \cdot q_{j+1}} \) are the same expression, in other words, for each permutation of J, the following holds:

\[
\frac{1}{1 - p_{j+1})(q_i - p_{j+1})(q_j - p_{j+1}) \cdot (q_{j+1} - p_{j+1})} - \frac{1}{q_i \cdot q_j \cdot q_{j+1}} = \left( \frac{1}{(1 - p_{j+1})(q_i - p_{j+1})(q_j - p_{j+1}) \cdot (q_{j+1} - p_{j+1})} + \frac{1}{q_i \cdot q_j \cdot q_{j+1}} \right)
\]

Which simplifies to:

\[
\frac{1}{(q_i - p_{j+1})(q_j - p_{j+1}) \cdot (q_{j+1} - p_{j+1})} - \frac{1}{q_i \cdot q_j \cdot q_{j+1}} = \left( \frac{1}{q_i \cdot q_j \cdot q_{j+1}} + \frac{1}{q_i \cdot q_j \cdot q_{j+1}} \right)
\]

Hence Lemma 1 holds if and only if:
For $n=1$, we obtain the identity:

$$q_i q_{i+1} \ldots q_j - (q_i - p_{j+1})(q_i - p_{j+1}) \ldots (q_j - p_{j+1}) =$$

$$p_{j+1} q_{i+1} q_{i+2} \ldots q_j q_i - q_i q_{i+1} q_{i+2} \ldots q_j =$$

Without loss of generality we simply note \( \{p_1, p_2, \ldots, p_j\} \) the permutation and consequently \( \{q_1, q_2, \ldots, q_j\} \) the intermediate variables.

Noting $X=p_{j+1}$ and with the convention that a product is equal to 1 if it is an empty product (i.e., the index range is empty) previous equality is equivalent to:

$$\prod_{i=1}^{j} q_i - \prod_{i=1}^{j} (q_i - X) = X \sum_{i=1}^{j} \left( \prod_{k=1}^{i-1} (q_k - X) \right) \left( \prod_{k=i+1}^{j} q_k \right).$$

It holds if and only if the polynomial coefficients (noted \([X^a]\) as usual) of both sides are equal \([X^a]RHS = [X^a]LHS\) for any $n$, $0 \leq n \leq j$.

This is obviously true for $n=0$ since: \( \prod_{i=1}^{j} q_i - \prod_{i=1}^{j} q_i = 0 \).

For $n=1$, we obtain the identity: \(- (\sum_{i=1}^{j} \prod_{k=1}^{i-1} q_k) = \sum_{i=1}^{j} \left( \prod_{k=1}^{i-1} q_k \right) \left( \prod_{k=i+1}^{j} q_k \right) \).

For the general case, one obtains on the left hand side:

$$[X^n]LHS = [X^n](- \prod_{i=1}^{j} (q_i - X)) = (-1)^{n+1} \sum_{J \subseteq \{1, \ldots, j\}} \prod_{i \in J} q_i$$

and on the right hand side:

$$RHS = \sum_{i=1}^{j} \left( (q_i - X) \prod_{k=1}^{i-1} (q_k - X) \prod_{k=i+1}^{j} q_k \right)$$

$$= - \sum_{i=1}^{j} \left( \prod_{k=1}^{i-1} (q_k - X) \prod_{k=i+1}^{j} q_k \right) + \sum_{i=1}^{j} \left( \prod_{k=1}^{i-1} (q_k - X) \prod_{k=i+1}^{j} q_k \right)$$

The $n$-th coefficient of the first sum is null for index $i < n$ and that of the second sum is null for $i-1 < n$ hence (noting $Q_J = \prod_{i \in J} q_i$ and $Q^i_J = \prod_{k=i}^{j} q_k$):

$$[X^n]RHS = - \sum_{i=n}^{j} (-1)^i \left( \sum_{J \subseteq \{i+1, \ldots, j\}} Q^i_J \right) Q_{i+1}^{i+1} + \sum_{i=n+1}^{j} (-1)^{i+1} \left( \sum_{J \subseteq \{i+1, \ldots, j\}} Q^i_J \right) Q_{i+1}^{i+1}.$$
\[[X^n]RHS = (-1)^{n+1} \left( \sum_{j=1}^{n} Q_j \right) Q_j^{j+1} \right) = [X^n]LHS \text{ since by convention } Q_j^{j+1} = 1.\]

This completes the proof of identity. QED.

**Appendix 2: Proof of Lemma 2**

Let \( J \) be a subset \( J \) of size \( j < m \) of the set \( \{1, \ldots, m\} \), \( j < m \), and \( F \) a numerical function. We want to prove that:

\[
\sum_{|J|=j} \sum_{i \in J} F(i, J) = \sum_{|J|=j+1} \sum_{i \in J} F(i, J - \{i\})
\]

For \( j = 1 \):

\[
\sum_{|J|=1} \sum_{i \in J} F(i, J) = \sum_{i=1}^{m} F(i, \{j\}) = \sum_{1 \leq j \leq m} F(i, \{j\}) + \sum_{1 \leq j \leq m} F(i, \{j\}) = \sum_{J = \{i, j\}} \sum_{i \in J} F(k, J - \{k\}).
\]

In the general case \( J = \{i_1, i_2, \ldots, i_j\} \), sum has \( j + 1 \) terms:

\[
\sum_{|J|=j} \sum_{k \in J} F(k, J) = \sum_{1 \leq i \leq j} F(i, J) + \sum_{1 \leq k \leq j} F(k, J) + \ldots + \sum_{1 \leq i \leq j} F(i, J) + \sum_{1 \leq k \leq j} F(k, J)
\]

Then denoting \( K \) the set \( \{i_1, i_2, \ldots, i_j, k\} \):

\[
\sum_{|J|=j+k} F(k, J) = \sum_{|K|=j+1} F(i_1, K - \{i_1\}) + \sum_{|K|=j+1} F(i_2, K - \{i_2\}) + \ldots + \sum_{|K|=j+1} F(i_j, K - \{i_j\}) + \sum_{|K|=j+1} F(k, K - \{k\})
\]

\[
= \sum_{|K|=j+1} \sum_{i \in K} F(i, K - \{i\}).
\]

QED.