Modern Stochastics: Theory and Applications 13 (3) (2026) 357–374
https://doi.org/10.15559/26-VMSTA298

Average precision at cutoff k under random rankings: expectation and variance

Tetiana Manzhos[a], **Tetiana Ianevych**[b,*], **Olga Melnyk**[a]

[a]*Kyiv National Economic University named after Vadym Hetman, 54/1 Beresteysky Avenue, 03057 Kyiv, Ukraine*
[b]*Taras Shevchenko National University of Kyiv, 60 Volodymyrska Street, 01033 Kyiv, Ukraine*

tmanzhos@kneu.edu.ua (T.V. Manzhos), tetianayanevych@knu.ua (T.O. Ianevych), melnyk.olha@kneu.edu.ua (O.O. Melnyk)

Received: 2 November 2025, Accepted: 28 February 2026, Published online: 2 April 2026

This paper is dedicated to the 85th anniversary of the birth of Prof. Yuriy Kozachenko.

Abstract Recommender systems and information retrieval platforms rely on ranking algorithms to present the most relevant items to users, thereby improving engagement and satisfaction. Assessing the quality of these rankings requires reliable evaluation metrics. Among them, Mean Average Precision at cutoff k (MAP@k) is widely used, as it accounts for both the relevance of items and their positions in the list for some groups of users.

It seems obvious that intelligent ranking algorithms should outperform recommendations generated at random. But how can we measure how much better they work? In this article, we have established the expected value and variance of the average accuracy at k (AP@k), as they can be used as a foundation for efficiency criteria for MAP@k. Here, we considered two widely used evaluation models: offline and online, together with corresponding randomization models for them, and calculated the expected value and variance of AP@k in both cases. The numerical study for different scenarios was also performed.

Keywords Recommendation systems, information retrieval, ranking algorithms, AP@k, MAP@k, evaluation metrics, expectation and variance, random rankings
2020 MSC 60E05, 60C05, 62R07, 68T05

*Corresponding author.

© 2026 The Author(s). Published by VTeX. Open access article under the CC BY license.

1 Introduction

Recommendation systems play a pivotal role in modern digital environments, designed to predict and suggest items such as movies on streaming platforms, products on e-commerce sites, or news articles on media platforms based on users' preferences and historical interactions. These systems leverage sophisticated algorithms that analyze user behavior, item attributes, and contextual data to provide personalized recommendations, thereby enhancing user satisfaction and engagement.

Ranking algorithms are fundamental to recommendation systems, as they determine the order in which items are presented to users. These algorithms prioritize relevant items and arrange them in a sequence designed to maximize user engagement and satisfaction.

Metrics that assess the effectiveness of these rankings are essential for evaluating algorithmic performance. Among various evaluation metrics, *Average Precision at k* (AP@k) plays a crucial role in recommendation systems due to its ability to assess the precision of recommended items up to a specified rank k. It not only accounts for the relevance of the items but also their position in the recommendation list, providing a comprehensive evaluation of ranking effectiveness. A review of this and other relevant metrics is provided in [7, 9], offering a deeper exploration of the evaluation methods.

AP@k is particularly important for evaluations of algorithms developed for scenarios where users are more likely to interact with top-ranked items, like book propositions on Amazon, short videos on YouTube, or news on Google News. While AP@k evaluates the ranking quality for individual users, MAP@k provides an overall measure of the algorithm's performance across the entire user base, giving a more generalized assessment of the system's ability to rank relevant items consistently.

There are a variety of different ranking methods, but none of them is the best. The effectiveness of a ranking algorithm depends on the particular problem, the data available, what rank must represent, and what properties are needed. The most well-known so far is the PageRank method used in search engines like Google (for details, see [4, 5] and the references therein).

In this paper, we calculated the closed-form expressions for the expected value and the variance of AP@k under random rankings. Since MAP@k averages user-level AP@k, the expectation of AP@k specifies the random baseline for the metric MAP@k, whereas the variance determines the scale of random fluctuations around that baseline. For a good ranking algorithm, MAP@k is expected to be significantly greater than the expectation of AP@k. This baseline supports principled benchmarking and guides the optimization of recommendation strategies. Together, these basic probabilistic parameters can be used for testing hypotheses on whether the algorithm is substantially better than just random ranking.

The paper is organized as follows. Section 2 introduces two evaluation settings that we consider (offline and online) in connection with corresponding randomization models and explains their practical relevance. Section 3 provides formal definitions of AP@k and MAP@k and clarifies their relationship. Sections 4 and 5 present closed-form expressions for the expectation and variance of AP@k in both offline and online scenarios. Section 6 presents some numerical examples and simulation results and comparative tables that illustrate the particularities of AP@k in several scenarios, whereas Section 7 summarizes the main findings and suggests directions for future

research. The proofs of the theorems from the Sections 4 and 5 are placed in the Appendix.

2 Offline and online evaluation

In the evaluation of recommendation systems, two principal approaches are widely used: *offline* evaluation and *online* one. A comprehensive discussion of these methods can be found in [1, 6]. In both these approaches, we restrict ourselves to treating item relevance as binary — classifying items as relevant or not — but they differ in the methodology.

It is important to note that the notion of relevance itself is not universal and is typically defined by the underlying business objective of the system. Depending on the task, relevance may correspond to different user actions, such as clicking on an item, visiting a product page, completing a purchase, or consuming a certain fraction of the content. For instance, in a recommendation system optimized for traffic generation, a click or page visit may be treated as a relevant outcome, whereas in a revenue-oriented setting, relevance may be defined by a completed transaction. Consequently, while AP@k and MAP@k assume a binary relevance indicator, the definition of this indicator is inherently task-dependent and reflects the specific optimization goal of the application.

Offline evaluation relies on historical data, where the relevance of items for each user is known. It is particularly useful when comprehensive, labeled data, such as historical interactions, is available, making it possible to test recommendation algorithms in a controlled environment. In this case, we can design a training set and a test set according to our preferences. The common practice is to group the users who have the same number of relevant items in the list in order to make them homogeneous in some sense and comparable. We can explore several groups with a different number m of relative items in the list.

What if we do not use the advanced ranking algorithms but propose the items to the user randomly? And what does it mean by 'randomly'? What will the expected value of AP@k be in this case? First of all, we need to define the random experiment that can be utilized under such a scenario. We can consider the experiment in which one random result differs from another only at the places where m relevant elements (no matter which) are located among N available places. It is stochastically equivalent to sampling m items from N with equal probabilities and without replacement (SWOR).

The offline evaluation is easier and reproducible. But it does not always predict online behavior and trends. Therefore, the offline evaluation is usually followed by the online evaluation. *Online* evaluation attempts to evaluate recommender systems by a method called A/B testing, where a part of users are served by recommender system A and another part of users by recommender system B in real-time mode, and then compare them with each other. We refer the interested reader to [1, 6] for more details on this issue. In A/B testing, one of the systems (i.e., B) can propose the items randomly. With time, all the recommender algorithms degrade, and the level of their degradation is regularly monitored, but instead of conducting an online test with random sorting of recommendations, which is quite costly and leads to reduced user satisfaction, it can be simulated using the randomization model.

We propose a model that also treats item relevance as binary. But it is unreasonable to require a predefined number of relevant items m within a set of N items per user, as was in the offline evaluation framework, since the model will be unrealistic. Instead, it is important to ensure that every participant in the sample testing the recommender system is shown at least k items from the recommendation list. For reliable evaluation in this scenario, two key conditions must be met: the pool of items used to generate the recommendation list must be sufficiently large, and the proportion of relevant items should remain approximately stable across all users in the treated group. In practice, it means that we need to choose users with similar characteristics, like gender or age. This comparison is facilitated by the probability p, which represents the likelihood of each item being relevant. The probability p can be estimated from the historical data. Given this probability, it is possible to calculate the expected value and the variance of the evaluation metric AP@k under random item ranking.

One possible random experiment that can be considered to fit this setting is Bernoulli sampling. It assumes that we have N items, and every item can be either relevant or not, independently of each other, with a probability of being relevant equal to p. This is equivalent to sampling from N items, where $m = pN$ (m is assumed to be natural) with equal probabilities and with replacement (SWR). The number of relevant items for each user under this model is not fixed, but on average it is equal to $Np = m$.

3 Average Precision at k and Mean Average Precision at k

Having established the context of offline and online evaluation, we now turn to the formal definition of the metrics Average Precision at k (AP@k) and Mean Average Precision at k (MAP@k), which are central to this paper. To begin, we first define AP@k for a single user, which serves as the building block for MAP@k.

The Average Precision at k measures the ranking quality of relevant items proposed to one user by averaging precision values at positions where relevant items appear. Formally, it is defined as:

$$AP@k = \frac{1}{k} \sum_{i=1}^{k} P@i \cdot I_i, \tag{1}$$

where:

- k is the length of the recommendation list;
- $P@i = \frac{\sum_{j=1}^{i} I_j}{i}$ is the precision at position i;
- $I_i \in \{0, 1\}$ is a binary indicator, equal to 1 if the item at position i is relevant and 0 otherwise.

In this general form, the denominator k normalizes the metric over the entire recommendation list shown to the user. However, when the total number of relevant items m is smaller than k, a modified normalization factor, $\min(m, k)$, should be used:

$$AP@k = \frac{1}{\min(m, k)} \sum_{i=1}^{k} P@i \cdot I_i. \tag{2}$$

This adjustment ensures that the metric appropriately handles situations where there are fewer relevant items than the length of the recommendation list, avoiding artificially deflated scores. Without this modification, even in the best scenario – where all m relevant items are ranked in the top positions – the AP@k score would never reach 1, because the denominator would be larger than m, preventing the score from reaching its maximum value.

As it was mentioned before, MAP@k aggregates the AP@k values across all users in the dataset and provides a more comprehensive evaluation of the recommendation system, thereby reflecting the system's overall performance of ranking. Formally, MAP@k is computed as:

$$MAP@k = \frac{1}{|U|} \sum_{u \in U} AP@k_u, \tag{3}$$

where:

- U is the set of all users in the dataset;
- $AP@k_u$ is the $AP@k$ score for user u.

Turning to the interpretation of the MAP@k metric, a high MAP@k value indicates that most users encounter relevant items ranked highly, reflecting an effective recommendation system. Conversely, a low MAP@k suggests that relevant items appear lower, signaling a less efficient algorithm. The metric ranges from 0 to 1, where 1 means all relevant items are within the top-k positions, and 0 means none are included. However, for a comprehensive evaluation, it is essential to compare the MAP@k score with the expected value derived from random recommendation generation. While one might assume that this expected value equals the proportion of relevant items in the collection, this is not the case. As demonstrated by Yves Bestgen in his work [2], the expected average precision for random recommendations is more complex and varies depending on various factors such as the number of relevant items in the list and other characteristics. In the current study, we derive precise formulas for the calculation of random baselines for both random ranking algorithms described above (sampling without and with replacement).

4 Expectation and variance of AP@k under sampling without replacement for the case of offline ranking evaluation

If we want to determine how much the algorithm outperforms random chance, we can compare MAP@k calculated for the developed ranking algorithm with the expected value of AP@k if randomness is imposed by sampling without replacement. The expected AP@k alone is not sufficient for a comprehensive evaluation, but it provides the crucial baseline against which meaningful improvements can be interpreted. In addition, the variance quantifies the spread of AP@k values around this baseline, completing the picture by indicating the scale of random fluctuations.

We begin with the offline evaluation setting described in Section 2, corresponding to sampling without replacement. In this case, among N items, exactly m are relevant, and one random outcome differs from another only by the positions of the m relevant

items in the ranking. Equivalently, the random ranking can be viewed as a uniformly random permutation of all N items, so that the m relevant items are placed at random positions. From the mathematical point of view, this means that for each item $j = 1, 2, \ldots, N$ we define a binary random variable I_j, with $I_j = 1$ if the item at position j is relevant and $I_j = 0$ otherwise. Furthermore, $\sum_{j=1}^{N} I_j = m$ and the random variables $I_1, I_2, \ldots, I_N$ are not independent.

Under this model, the probability that the j-th item is relevant can be calculated as

$$P(I_j = 1) = \frac{C_{N-1}^{m-1}}{C_N^m} = \frac{m}{N}.$$

So, $E(I_j) = m/N$ for every j and the random variable $\sum_{j=1}^{i} I_j$, which counts the number of relevant items among the first i items, follows the *Hypergeometric* distribution with parameters N, m and i. Moreover

$$\mathrm{E}(P@i) = \frac{1}{i}\sum_{j=1}^{i} \mathrm{E}(I_j) = \frac{m}{N}, \quad i = 1, \ldots, k.$$

In this case, we should compute $AP@k$ according to formula (2), where the normalization is made by $\min(m, k)$.

Hereafter, for the identification that the expectations are calculated with respect to different random models, we used different subscripts (SWOR and SWR) in the theorems' formulation. In the proofs placed in Appendix A, we omit these subscripts to simplify the formulas.

The explicit expressions for the expected value $\mathrm{E}(AP@k)$ and the variance $\mathrm{Var}(AP@k)$ under the random model corresponding to sampling without replacement (SWOR) are established in the following theorem.

Theorem 1. *Assume that among N available items, exactly m are relevant for each user and their places are chosen randomly with equal probability by sampling without replacement. Then the expectation and the variance of AP@k with respect to all possible random results are equal to*

$$\mathrm{E}_{SWOR}(AP@k) = \frac{m}{N \cdot \min(k, m)} \left(\frac{m-1}{N-1} k + \frac{N-m}{N-1} H_k \right), \tag{4}$$

$$\mathrm{Var}_{SWOR}\left(AP@k\right) = \frac{1}{\min(k, m)^2} \frac{m}{N} \Bigg[k(C + 2(E - F) + (k-1)G) + \\ + H_k(B - 2(E - kF)) + H_k^2 \cdot D + H_k^{(2)}(A - D) \Bigg],$$

where

$$A = 1 - \frac{m}{N} - \frac{m-1}{N-1}\left(3 - 2\frac{m-2}{N-2} - \frac{m}{N}\left(2 - \frac{m-1}{N-1}\right)\right);$$

$$B = \frac{m-1}{N-1}\left(3\left(1 - \frac{m-2}{N-2}\right) - 2\frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)\right);$$

$$C = \frac{m-1}{N-1}\left(\frac{m-2}{N-2} - \frac{m(m-1)}{N(N-1)}\right);$$

$$D = \frac{m-1}{N-1}\left(2 - 5\frac{m-2}{N-2} + 3\frac{(m-2)(m-3)}{(N-2)(N-3)}\right) - \frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)^2;$$

$$E = \frac{m-1}{N-1}\left(3\frac{m-2}{N-2}\left(1 - \frac{m-3}{N-3}\right) - \frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)\right);$$

$$F = \frac{m-1}{N-1}\left(\frac{m-2}{N-2}\left(1 - \frac{m-3}{N-3}\right) - \frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)\right);$$

$$G = \frac{m-1}{N-1}\left(\frac{(m-2)(m-3)}{(N-2)(N-3)} - \frac{m}{N}\frac{m-1}{N-1}\right);$$

$$H_k = \sum_{i=1}^{k}\frac{1}{i}; \quad H_k^{(2)} = \sum_{i=1}^{k}\frac{1}{i^2}.$$

Remark 1. The values $H_k = \sum_{i=1}^{k}\frac{1}{i}$ are, actually, the partial sums of the harmonic series. They were named 'harmonic' numbers in 1968 by Donald Knuth [8]. For quite big k, the well-known formula (see [3])

$$H_k = \ln k + \gamma + \frac{1}{2k} - \varepsilon_k,$$

can be used. Here $\gamma \approx 0.5772$ is the Euler–Mascheroni constant and $0 \le \varepsilon_k \le 1/8k^2$ which tends to 0 as $k \to \infty$.

Remark 2. For small k, the values H_k and $H_k^{(2)}$ can be calculated exactly by ordinary summation.

5 Expectation and variance of AP@k under sampling with replacement in case of the online ranking evaluation

We now turn to the online evaluation setting described in Section 2. Here, we assume that each item is relevant to a user with a constant probability p. The randomness can be modeled as a sequence of independent Bernoulli trials, or equivalently, as sampling with replacement (SWR), where each item has probability p of being sampled. In this case, the decision whether a recommendation is relevant is assumed to be independent for each item. Moreover, we assume a uniform relevance probability p for all recommendations and all users. These probabilistic assumptions differ from the offline case and lead to distinct formulas for the expectation and variance of AP@k.

From a mathematical point of view, this means that for each recommendation $j = 1, 2, \ldots, N$, we define the random variables $I_j = 1$ if it is relevant and $I_j = 0$ otherwise, as in the previous section. But here the random variables $I_1, I_2, \ldots, I_N$ are independent with mean $\mathrm{E}(I_j) = p$ and variance $\mathrm{Var}(I_j) = p(1-p)$, where p is assumed to be known. Of course, this assumption is quite strong and can be applied only when the users are relatively similar (homogeneous), as described in Section 2 for the online evaluation scenario.

Under these assumptions and notations, we have the following:

$$P@i = \frac{1}{i}\sum_{j=1}^{i} I_j, \quad \mathrm{E}(P@i) = \frac{1}{i}\sum_{j=1}^{i}\mathrm{E}(I_j) = p, \quad \text{and} \quad AP@k = \frac{1}{k}\sum_{i=1}^{k} P@i \cdot I_i.$$

The basic benchmarking characteristics for $MAP@k$ under random recommendations corresponding to the sampling with replacement are presented in the following theorem.

Theorem 2. *If, for each user, every item is assumed to be independently relevant with constant probability p and the places for the relevant items in the list of N items are chosen using sampling with replacement with probability p, then the expectation and the variance of AP@k are equal to*

$$\mathrm{E}_{SWR}(AP@k) = p\left(p + (1-p)\frac{1}{k}H_k\right); \tag{5}$$

$$\mathrm{Var}_{SWR}(AP@k) = \frac{5}{k}\cdot p^3(1-p) + \frac{1}{k^2}\cdot p(1-p)\left[p(1-2p)(3H_k + H_k^2) + (1-p)(1-3p)H_k^{(2)}\right].$$

Here $H_k = \sum_{i=1}^{k}\frac{1}{i}$ is the k-th harmonic number and $H_k^{(2)} = \sum_{i=1}^{k}\frac{1}{i^2}$ is the k-th partial sum of the 2nd order harmonic series as well as in Theorem 1.

6 Numerical examples and simulations results

Before presenting some practical results, it is helpful to summarize the key differences in the formulas for the offline and online settings. In the *offline* case, the distribution of AP@k depends on the total number of items N, the number of relevant items m, and the cut-off k that specifies how many top-ranked positions are taken into account. This reflects the fact that evaluation is performed on a finite set of items, where exactly m out of N are relevant and their positions in the ranking determine the outcome. In the *online* setting, by contrast, the parameter N does not appear in the formulas, since evaluation is restricted to the top-k recommendations and each of them is modeled as an independent Bernoulli trial with relevance probability p. This makes the outcome depend only on k and p, regardless of the overall size of the candidate pool, which in practice may be very large. Conceptually, the probability p plays a role analogous to the ratio m/N in the offline case: while offline evaluation fixes the proportion of relevant items deterministically, online evaluation assumes it probabilistically.

To cover a representative range of conditions, $N = 50$ is fixed, and several choices of m and k are examined. The half-density case ($m = 25$) with varying cut-offs illustrates the situations $m > k$, $m = k$, and $m < k$. Lower prevalences ($m = 10$ and $m = 2$) are also included along with a high-prevalence setting ($m = 35$, corresponding to $p > 0.5$). These scenarios span the spectrum from very sparse to very dense relevance distributions, thereby testing the formulas under diverse conditions. The selected configurations are summarized in Table 1.

Table 1. Examined scenarios

Scenario	Offline: (N, m)	Online: p	k	Characteristic
A1	(50, 25)	0.50	5	$m > k$, balanced case (50% relevant)
A2	(50, 25)	0.50	25	$m = k$, balanced case (50% relevant)
A3	(50, 25)	0.50	40	$m < k$, balanced case (50% relevant)
B	(50, 10)	0.20	20	moderate proportion of relevant items
C	(50, 2)	0.04	20	very low proportion of relevant items
D	(50, 35)	0.70	20	high proportion of relevant items ($p > 0.5$)

Table 2 reports the theoretical (derived from the formulas in Sections 4 and 5) means and variances of AP@k for the selected scenarios. The offline and online settings are placed side by side, allowing examination of their behavior under comparable conditions and observation of where the two randomization schemes produce similar outcomes and where differences emerge.

Table 2. Values of expectation and variance for AP@k for scenarios from Table 1

Scenario	SWOR Expectation	SWR Expectation	SWOR Variance	SWR Variance
A1	0.36139	0.36416	0.05464	0.05884
A2	0.28387	0.28816	0.00735	0.01234
A3	0.43550	0.27674	0.00699	0.00775
B	0.13221	0.06878	0.00786	0.00294
C	0.07865	0.00851	0.01563	0.00023
D	0.52426	0.52778	0.01502	0.02195

As seen in Table 2, the comparison between offline (SWOR) and online (SWR) shows no uniform pattern: in some scenarios the two models behave very similarly, while in others their outcomes diverge noticeably. The observed differences reflect the $\min(m, k)$ normalization in offline evaluation and the independence structure of online sampling. Together, these results provide a clear basis for interpreting MAP@k under the random rankings.

In addition to the values reported in Table 2, Figures 1 and 2 show the simulated distributions of AP@k for two representative scenarios. These examples illustrate how

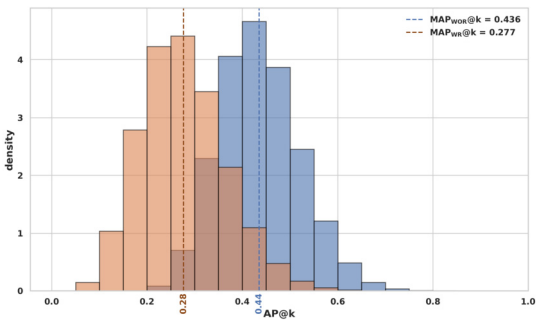


Fig. 1. Histogram of AP@k values for Scenario A3 ($N = 50, m = 25, p = 0.5, k = 40$). The SWOR distribution (blue) is shifted to higher values compared to SWR (orange), reflecting normalization by $\min(m, k)$ in the offline case

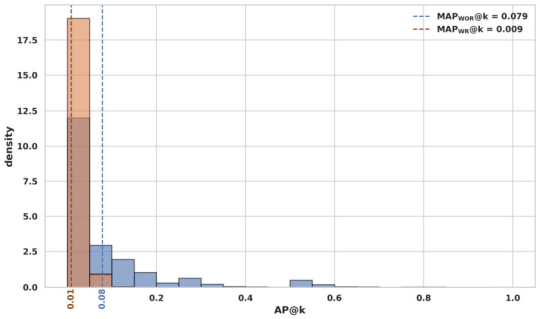


Fig. 2. Histogram of AP@k values for Scenario C ($N = 50, m = 2, p = 0.04, k = 20$). The SWR distribution (orange) is tightly concentrated around zero, while the SWOR distribution (blue) shows wider variability due to the finite number of relevant items

SWOR and SWR can differ not only in their means but also in their variability.

In *Scenario A3* ($m < k$), the distributions are shifted relative to each other: the offline case yields systematically higher values due to normalization by $\min(m, k)$, while the online case, which models relevance probabilistically, results in lower means.

In *Scenario C* (very low prevalence), the contrast lies in variability. The SWR distribution is sharply concentrated, reflecting that almost always no relevant items appear in the top-k, whereas the SWOR distribution remains wider because the fixed number of relevant items can still occupy different positions in the ranking.

7 Conclusions and discussion

The results obtained clarify what levels of AP@k are expected under random ranking and how much variability surrounds them. From a practical standpoint, these findings offer two key insights. First, the expected value of AP@k serves as a clear reference point against which observed performance can be compared. Second, the variance quantifies the random fluctuations around this reference, helping to distinguish genuine improvements from chance outcomes. Together, expectation and variance can be used to build the statistical test to check the hypothesis on whether the studied new ranking algorithm is significantly better than a random one or not by comparing the empirical MAP@k obtained from offline or online evaluation procedure with $\mathrm{E}(AP@k)$. But this task is left for future research. It will also be interesting to extend this analysis to other related metrics and investigate how statistical testing frameworks can build on these results for a robust comparison of ranking systems.

A Appendix. Proofs of Theorem 1 and 2

In the proofs, we need some technical results, which we formulate as a lemma.

Lemma 1. *For any* $k \geq 1$

(i) $\sum_{i=1}^{k-1} \sum_{l=i+1}^{k} \frac{1}{i} = k(H_k - 1)$;

(ii) $\sum_{i=1}^{k-1} \sum_{l=i+1}^{k} \frac{1}{l} = k - \sum_{l=1}^{k} \frac{1}{l} = k - H_k$;

(iii) $\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\frac{1}{i\cdot l}=\frac{1}{2}\left(H_k^2-H_k^{(2)}\right)$,

where $H_k=\sum_{i=1}^{k}\frac{1}{i}$ *is a standard harmonic number;* $H_k^{(2)}=\sum_{i=1}^{k}\frac{1}{i^2}$ *is the k-th partial sum of the 2nd order harmonic series.*

Proof. (i)

$$\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\frac{1}{i}=\sum_{i=1}^{k-1}\frac{1}{i}(k-i)=k\sum_{i=1}^{k-1}\frac{1}{i}-(k-1)+(\frac{1}{k}-\frac{1}{k})k=k\sum_{i=1}^{k}\frac{1}{i}-k=k(H_k-1).$$

(ii)

$$\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\frac{1}{l}=\sum_{l=2}^{k}\sum_{i=1}^{l-1}\frac{1}{l}=\sum_{l=2}^{k}\frac{l-1}{l}=k-1-\left(\sum_{l=1}^{k}\frac{1}{l}-1\right)=k-\sum_{l=1}^{k}\frac{1}{l}=k-H_k.$$

(iii) For the calculation of the last equality, we can use the fact that for any finite collection of numbers $\{a_{il}\}_{i,l=1}^{k}$, such that $a_{il}=a_{li}$ if $i\neq l$

$$\sum_{i=1}^{k-1}\sum_{l=1+1}^{k}a_{il}=\frac{1}{2}\left(\sum_{i=1}^{k}\sum_{l=1}^{k}a_{il}-\sum_{i=1}^{k}a_{ii}\right).$$

So,

$$\begin{aligned}\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\frac{1}{i\cdot l}&=\frac{1}{2}\left(\sum_{i=1}^{k}\sum_{l=i+1}^{k}\frac{1}{i\cdot l}-\sum_{i=1}^{k}\frac{1}{i^2}\right)\\&=\frac{1}{2}\left(\left(\sum_{i=1}^{k}\frac{1}{i}\right)^2-\sum_{i=1}^{k}\frac{1}{i^2}\right)=\frac{1}{2}\left(H_k^2-H_k^{(2)}\right).\end{aligned}$$

□

Proof of Theorem 1. *For the expectation.*

$$\mathrm{E}(AP@k)=\frac{1}{\min(m,k)}\sum_{i=1}^{k}\mathrm{E}(P@i\cdot I_i).$$

Let's first calculate the following:

$$\begin{aligned}\mathrm{E}(P@i\cdot I_i)&=\mathrm{E}\left(\mathrm{E}(P@i\cdot I_i|I_i)\right)=\\&=\mathrm{E}\left(P@i\cdot I_i|I_i=1\right)P\{I_i=1\}+\mathrm{E}\left(P@i\cdot I_i|I_i=0\right)P\{I_i=0\}=\\&=\mathrm{E}\left(P@i|I_i=1\right)\cdot\frac{m}{N}=\frac{1}{i}\left((i-1)\frac{m-1}{N-1}+1\right)\cdot\frac{m}{N},\end{aligned}$$

since, if we know that $I_i=1$, then $P@i$ is distributed as $\frac{1}{i}\cdot\left(Hypergeometric(N-1,m-1,i-1)+1\right)$. Therefore,

$$\begin{aligned}\mathrm{E}(AP@k)&=\frac{1}{\min(m,k)}\cdot\frac{m}{N}\sum_{i=1}^{k}\left((1-\frac{1}{i})\frac{m-1}{N-1}+\frac{1}{i}\right)=\\&=\frac{1}{\min(m,k)}\cdot\frac{m}{N}\left(k\cdot\frac{m-1}{N-1}+(1-\frac{m-1}{N-1})\sum_{i=1}^{k}\frac{1}{i}\right).\end{aligned}$$

For the variance.

Since $\mathrm{Var}(AP@k) = \mathrm{Var}\left(\frac{1}{k}\sum_{i=1}^{k} P@i \cdot I_i\right) = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{l=1}^{k} c_{il}$, we need to calculate all the coefficient c_{il} taking into account that for every $i \neq l$ $c_{il} = c_{li} = \mathrm{cov}(P@i \cdot I_i, P@l \cdot I_l)$ and for all i $c_{ii} = \mathrm{Var}(P@i \cdot I_i)$.

Let us calculate them step by step, starting from the coefficients $c_{ii}, i = 1, 2, \ldots, k$. Since in this case the random variables $I_1, \ldots, I_n$ are not independent, it is worth applying formulas that use conditional expectations. In particular, for two random variables X and Y, we can calculate the variance by the formula:

$$\mathrm{Var}(X) = \mathrm{E}\left(\mathrm{Var}(X \mid Y)\right) + Var\left(\mathrm{E}(X \mid Y)\right).$$

So,

$$c_{ii} = \mathrm{Var}\left(P@i \cdot I_i\right) = \mathrm{E}\left(\mathrm{Var}(P@i \cdot I_i \mid I_i\right) + \mathrm{Var}\left(\mathrm{E}(P@i \cdot I_i \mid I_i)\right) = t_1 + t_2. \quad (6)$$

If $I_i = 0$, then $P@i \cdot I_i = 0$, and $\mathrm{E}\left(P@i \cdot I_i \mid I_i = 0\right) = 0$, $\mathrm{Var}\left(P@i \cdot I_i \mid I_i = 0\right) = 0$. If $I_i = 1$, then $\mathrm{E}\left(P@i \cdot I_i \mid I_i = 1\right) = \frac{1}{i}\left((i-1)\frac{m-1}{N-1} + 1\right)$ and

$$\mathrm{Var}\left(P@i \cdot I_i \mid I_i = 1\right) = \mathrm{E}\left(\left(P@i \cdot I_i\right)^2 \mid I_i = 1\right) - \left(\mathrm{E}(P@i \cdot I_i \mid I_i = 1)\right)^2.$$

First of all

$$\begin{aligned}
\mathrm{E}\left((P@i \cdot I_i)^2 \mid I_i = 1\right) &= \mathrm{E}\left(\left(\frac{1}{i}\left(\sum_{j=1}^{i-1} I_j I_i + I_i^2\right)\right)^2 \mid I_i = 1\right) = \\
&= \frac{1}{i^2}\,\mathrm{E}\left(1 + 2\sum_{j=1}^{i-1} I_j + \left(\sum_{j=1}^{i-1} I_j\right)^2 \mid I_i = 1\right) = \\
&= \frac{1}{i^2}\,\mathrm{E}\left(1 + 2\sum_{j=1}^{i-1} I_j + \sum_{j=1}^{i-1} I_j^2 + \sum\sum_{j\neq r} I_j I_r \mid I_i = 1\right) = \\
&= \frac{1}{i^2}\left(1 + 3\sum_{j=1}^{i-1} \mathrm{E}\left(I_j \mid I_i = 1\right) + \sum\sum_{j\neq r} \mathrm{E}\left(I_j I_r \mid I_i = 1\right)\right) = \\
&= \frac{1}{i^2}\left(1 + 3(i-1)\cdot\frac{m-1}{N-1} + (i-1)(i-2)\frac{(m-1)(m-2)}{(N-1)(N-2)}\right),
\end{aligned}$$

as $\mathrm{E}\left(I_j \mid I_i = 1\right) = \frac{C_{N-2}^{m-2}}{C_{N-1}^{m-1}} = \frac{m-1}{N-1}$ and $\mathrm{E}(I_j I_r \mid I_i = 1) = \frac{C_{N-3}^{m-3}}{C_{N-1}^{m-1}} = \frac{(m-1)(m-2)}{(N-1)(N-2)}$.

Then

$$\begin{aligned}
\mathrm{Var}(P@i \cdot I_i \mid I_i = 1) = &\frac{1}{i^2}\left(1 + 3(i-1)\cdot\frac{m-1}{N-1} + (i-1)(i-2)\frac{(m-1)(m-2)}{(N-1)(N-2)}\right) \\
&- \frac{1}{i^2}\left((i-1)\frac{m-1}{N-1} + 1\right)^2
\end{aligned}$$

and thus we obtain the first term in (6):

$$t_1 = \mathrm{E}\left(\mathrm{Var}(P@i \cdot I_i \mid I_i)\right) = \mathrm{Var}\left(P@i \cdot I_i \mid I_i = 1\right) \cdot \frac{m}{N} =$$
$$= \frac{m}{N} \cdot \frac{1}{i^2}\left(1 + 3(i-1) \cdot \frac{m-1}{N-1} + (i-1)(i-2)\frac{(m-1)(m-2)}{(N-1)(N-2)}\right.$$
$$\left. -((i-1)\frac{m-1}{N-1} + 1)^2\right).$$

Secondly, if $I_i = 1$ the random variable $\mathrm{E}(P@i \cdot I_i \mid I_i)$ takes the value $\frac{1}{i} \cdot \left((i-1)\frac{m-1}{N-1} + 1\right)$ with probability $\frac{m}{N}$ and 0 otherwise. Therefore,

$$t_2 = \mathrm{Var}\left(\mathrm{E}(P@i \cdot I_i \mid I_i)\right) = \mathrm{E}\left((\mathrm{E}(P@i \cdot I_i \mid I_i))^2\right) - \left(\mathrm{E}(\mathrm{E}(P@i \cdot I_i \mid I_i))\right)^2 =$$
$$= \frac{1}{i^2}\left((i-1)\frac{m-1}{N-1} + 1\right)^2 \cdot \frac{m}{N} - \frac{1}{i^2}\left((i-1)\frac{m-1}{N-1} + 1\right)^2 \cdot \left(\frac{m}{N}\right)^2 =$$
$$= \frac{1}{i^2}\left((i-1)\frac{m-1}{N-1} + 1\right)^2 \cdot \frac{m}{N}\left(1 - \frac{m}{N}\right).$$

So, finally, we have the following expression after collecting the coefficients near $\frac{1}{i^r}, r = 0, 1, 2$.

$$c_{ii} = \mathrm{Var}\left(P@i \cdot I_i\right) = t_1 + t_2 =$$
$$= \frac{m}{N} \cdot \frac{1}{i^2}\left(1 + 3(i-1) \cdot \frac{m-1}{N-1} + (i-1)(i-2)\frac{(m-1)(m-2)}{(N-1)(N-2)}\right) -$$
$$- \frac{m}{N} \cdot \frac{1}{i^2}\left((i-1)\frac{m-1}{N-1} + 1\right)^2 + \frac{1}{i^2}\left((i-1)\frac{m-1}{N-1} + 1\right)^2 \cdot \frac{m}{N}\left(1 - \frac{m}{N}\right) =$$
$$= \frac{m}{N}\left(1 - \frac{m}{N} - \frac{m-1}{N-1}\left(3 - 2\frac{m-2}{N-2} - \frac{m}{N}\left(2 - \frac{m-1}{N-1}\right)\right)\right)\frac{1}{i^2} +$$
$$+ \frac{m}{N}\frac{(m-1)}{(N-1)}\left(3\left(1 - \frac{m-2}{N-2}\right) - 2\frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)\right)\frac{1}{i} +$$
$$+ \frac{m}{N}\frac{(m-1)}{(N-1)}\frac{(m-2)}{(N-2)} - \left(\frac{m(m-1)}{N(N-1)}\right)^2.$$

(ii) For all such i and l that $1 < i < l \leq k$

$$c_{il} = c_{li} = \mathrm{cov}(P@i \cdot I_i, P@l \cdot I_l) = \mathrm{E}(P@i \cdot I_i \cdot P@l \cdot I_l) - \mathrm{E}(P@i \cdot I_i) \cdot \mathrm{E}(P@l \cdot I_l).$$

Here, we can also use the conditional expectation for calculation

$$\mathrm{E}(P@iI_i \cdot P@lI_l) = \mathrm{E}\left(\mathrm{E}(P@iI_i \cdot P@lI_l \mid I_iI_l)\right)$$
$$= \mathrm{E}(P@iI_i \cdot P@lI_l \mid I_iI_l = 1)P(I_iI_l = 1).$$

If $I_iI_l = 1$, then we know that the items in the ith and lth places are relevant, so we should take into account the possible arrangements of the other $m - 2$ relevant items

between the $N-2$ places. So, $P(I_i I_l = 1) = \frac{C_{N-2}^{m-2}}{C_N^m} = \frac{m(m-1)}{N(N-1)}$. And then

$$\begin{aligned}
\mathrm{E}(P@i \cdot I_i \cdot P@l \cdot I_l \mid I_i I_l = 1) &= \mathrm{E}\left(\frac{1}{i}\sum_{j=1}^{i} I_j \cdot \frac{1}{l}\sum_{r=1}^{l} I_r \mid I_i I_l = 1\right) \\
&= \frac{1}{i \cdot l}\,\mathrm{E}\left(\left(1 + \sum_{j=1}^{i-1} I_j\right)\left(1 + \sum_{r=1}^{i-1} I_r + \sum_{r=i+1}^{l-1} I_r + 1\right) \mid I_i I_l = 1\right) \\
&= \frac{1}{i \cdot l}\,\mathrm{E}\left(2 + 4\sum_{r=1}^{i-1} I_r + \sum_{r=i+1}^{l-1} I_r + 2\sum_{j=1}^{i-r}\sum_{r=j+1}^{i-1} I_j I_r + \sum_{j=1}^{i-1}\sum_{r=i+1}^{l-1} I_j I_r \mid I_i I_l = 1\right) \\
&= \frac{1}{i \cdot l}\left(2 + 4(i-1)\frac{m-2}{N-2} + (l-1-i)\frac{m-2}{N-2} + (i-1)(i-2)\frac{(m-2)(m-3)}{(N-2)(N-3)}\right. \\
&\quad \left. + (i-1)(l-1-i)\frac{(m-2)(m-3)}{(N-2)(N-3)}\right) = \\
&= \frac{1}{i \cdot l}\left(2 + \frac{m-2}{N-2}(3i-5+l) + \frac{(m-2)(m-3)}{(N-2)(N-3)}(i-1)(l-3)\right).
\end{aligned}$$

So, after putting all the components together and doing some algebra, we obtain

$$\begin{aligned}
c_{il} &= \frac{m(m-1)}{N(N-1)}\frac{1}{i \cdot l}\left[2 - 5\frac{m-2}{N-2} + i \cdot 3\frac{m-2}{N-2} + l\frac{m-2}{N-2} + i \cdot l\frac{(m-2)(m-3)}{(N-2)(N-3)} - \right. \\
&\quad \left. - i \cdot 3\frac{(m-2)(m-3)}{(N-2)(N-3)} - l\frac{(m-2)(m-3)}{(N-2)(N-3)} + 3\frac{(m-2)(m-3)}{(N-2)(N-3)}\right] - \\
&\quad - \left(\frac{m}{N}\right)^2 \frac{1}{i \cdot l}\left(1 + (i-1)\frac{m-1}{N-1}\right)\left(1 + (l-1)\frac{m-1}{N-1}\right) = \\
&= \frac{m}{N}\frac{1}{i \cdot l}\left[\frac{m-1}{N-1}\left(2 - 5\frac{m-2}{N-2} + 3\frac{(m-2)(m-3)}{(N-2)(N-3)}\right) - \frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)^2 + \right. \\
&\quad + i\frac{m-1}{N-1}\left(3\frac{m-2}{N-2}\left(1 - \frac{m-3}{N-3}\right) - \frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)\right) + \\
&\quad + l\frac{m-1}{N-1}\left(\frac{m-2}{N-2}\left(1 - \frac{m-3}{N-3}\right) - \frac{m}{N}\left(1 - \frac{m-1}{N-1}\right)\right) + \\
&\quad \left. + i \cdot l\frac{m-1}{N-1}\left(\frac{(m-2)(m-3)}{(N-2)(N-3)} - \frac{m}{N}\frac{m-1}{N-1}\right)\right].
\end{aligned}$$

Using the notations from the Theorem 1 statement and $M = min(m, k)$ for shortness, along with results from Lemma 1, we'll get:

$$\begin{aligned}
\mathrm{Var}_{SWOR}(AP@k) &= \frac{1}{M^2}\left(\sum_{i=1}^{k} c_{ii} + 2\sum_{i=1}^{k-1}\sum_{l=i+1}^{k} c_{il}\right) = \\
&= \frac{1}{M^2}\left[\sum_{i=1}^{k}\frac{m}{N}\left(\frac{A}{i^2} + \frac{B}{i} + C\right) + 2\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\frac{m}{N}\left(\frac{D}{i \cdot l} + \frac{E}{l} + \frac{F}{i} + G\right)\right] =
\end{aligned}$$

$$=\frac{1}{M^2}\frac{m}{N}\Bigg[A\cdot H_k^{(2)}+B\cdot H_k+kC+D(H_k^2-H_k^{(2)})+$$

$$+2Ek(H_k-1)+2F(k-H_k)+Gk(k-1)\Bigg]=$$

$$=\frac{1}{M^2}\frac{m}{N}\Bigg[k[C+2(E-F)+(k-1)G]+H_k(B-2(E-kF))+$$

$$+H_k^2\cdot D+H_k^{(2)}(A-D)\Bigg].$$

This completes the proof. □

Proof of Theorem 2. *For the expectation.*

If we examine the model of relevance of proposals that corresponds to sampling with replacement with a fixed probability of success p, then we can use $\frac{1}{k}$ for normalization purposes, since in this case it is possible that all k items can be relevant. So, by the definition of $AP@k$, we have

$$\mathrm{E}(AP@k)=\mathrm{E}\left(\frac{1}{k}\sum_{i=1}^{k}P@i\cdot I_i\right)=\frac{1}{k}\sum_{i=1}^{k}\mathrm{E}\left(\frac{1}{i}\sum_{j=1}^{i}I_j\cdot I_i\right)$$
$$=\frac{1}{k}\sum_{i=1}^{k}\left(\frac{1}{i}\sum_{j=1}^{i}\mathrm{E}(I_j\cdot I_i)\right).$$

For all $j\neq i$, $\mathrm{E}(I_j\cdot I_i)=\mathrm{E}(I_j)\cdot\mathrm{E}(I_i)=p^2$, as these random variables are independent. If $j=i$, then $\mathrm{E}(I_i^2)=\mathrm{E}(I_i)=P\{I_i=1\}=p$. So,

$$\mathrm{E}(AP@k)=\frac{1}{k}\sum_{i=1}^{k}\frac{1}{i}\left(\sum_{j=1}^{i-1}\mathrm{E}(I_j\cdot I_i)+\mathrm{E}(I_i)\right)=\frac{1}{k}\sum_{i=1}^{k}\frac{1}{i}((i-1)p^2+p)=$$
$$=\frac{p}{k}\sum_{i=1}^{k}(p+(1-p)\frac{1}{i})=p^2+p(1-p)\frac{1}{k}\sum_{i=1}^{k}\frac{1}{i}.$$

For the variance. As before,

$$\mathrm{Var}(AP@k)=\mathrm{Var}\left(\frac{1}{k}\sum_{i=1}^{k}P@i\cdot I_i\right)=\frac{1}{k^2}\sum_{i=1}^{k}\sum_{l=1}^{k}c_{il},$$

and we need to calculate the coefficient c_{il} taking into account that for every $i\neq l$ $c_{il}=c_{li}=\mathrm{cov}(P@i\cdot I_i,P@l\cdot I_l)$ and for all i $c_{ii}=\mathrm{Var}(P@i\cdot I_i)$.

Let us calculate them again one by one.

(i) For $i = 1, 2, \ldots, k$

$$c_{ii} = \mathrm{Var}\left(\frac{1}{i}\sum_{j=1}^{i} I_j I_i\right) = \frac{1}{i^2}\,\mathrm{Var}\left(\sum_{j=1}^{i-1} I_j I_i + I_i^2\right) =$$
$$= \frac{1}{i^2}\,\mathrm{Var}\left(I_i\left(1 + \sum_{j=1}^{i-1} I_j\right)\right).$$

Here, the random variables I_i and $1 + \sum_{j=1}^{i-1} I_j$ are independent. The variance of the product of two independent random variables X and Y can be calculated using the following rule:

$$\mathrm{Var}(X \cdot Y) = \mathrm{E}(X^2)\,\mathrm{E}\left(Y^2\right) - \left(\mathrm{E}(X)\right)^2\left(\mathrm{E}(Y)\right)^2.$$

As far as $\mathrm{E}\left(I_i^2\right) = \mathrm{E}(I_i) = p$,
$\mathrm{E}\left(1 + \sum_{j=1}^{i-1} I_j\right) = 1 + \sum_{j=1}^{i-1} \mathrm{E}\left(I_j\right) = 1 + (i-1)p$, and

$$\mathrm{E}\left(\left(1 + \sum_{j=1}^{i-1} I_j\right)^2\right) = \mathrm{E}\left(1 + 2\sum_{j=1}^{i-1} I_j + \left(\sum_{j=1}^{i-1} I_j\right)^2\right) =$$
$$= \mathrm{E}\left(1 + 2\sum_{j=1}^{i-1} I_j + \sum_{j=1}^{i-1} I_j^2 + 2\sum_{j=1}^{i-2}\sum_{r=j+1}^{i-1} I_j I_r\right) =$$
$$= 1 + 3(i-1)p + (i-1)(i-2)p^2,$$

then

$$c_{ii} = \frac{1}{i^2}\,\mathrm{Var}\left(I_i \cdot \left(1 + \sum_{j=1}^{i-1} I_j\right)\right) =$$
$$= \frac{1}{i^2}\left(\mathrm{E}\left(I_i^2\right)\mathrm{E}\left(\left(1 + \sum_{j=1}^{i-1} I_j\right)^2\right) - (\mathrm{E}(I_i))^2\left(\mathrm{E}\left(1 + \sum_{j=1}^{i-1} I_j\right)\right)^2\right) =$$
$$= \frac{1}{i^2}\left(p\left(1 + 3(i-1)p + (i-1)(i-2)p^2\right) - p^2\left(1 + (i-1)p\right)^2\right) =$$
$$= p(1-p)\left(p^2 + p(3-2p)\frac{1}{i} + (1 - 3p + p^2)\frac{1}{i^2}\right).$$

(ii) For all i and l for which $1 < i < l \le k$

$$c_{il} = c_{li} = \mathrm{cov}(I_i \cdot P@i, I_l \cdot P@l) = \mathrm{E}(I_i \cdot P@i \cdot I_l \cdot P@l) - \mathrm{E}(I_i \cdot P@i) \cdot \mathrm{E}(I_l \cdot P@l).$$

Here

$$\begin{aligned}
\mathrm{E}(I_i \cdot P@i \cdot I_l \cdot P@l) &= \frac{1}{i \cdot l}\,\mathrm{E}\left(I_i\left(1+\sum_{j=1}^{i-1} I_j\right)\cdot I_l\left(1+\sum_{r=1}^{l-1} I_r\right)\right)\\
&= \frac{1}{i \cdot l}\,\mathrm{E}\left(I_i I_l\left(1+\sum_{j=1}^{i-1} I_j\right)\left(1+\sum_{r=1}^{i-1} I_r + I_i + \sum_{r=i+1}^{l-1} I_r\right)\right)\\
&= \frac{1}{i \cdot l}\,\mathrm{E}\left(I_i I_l\left(1+\sum_{j=1}^{i-1} I_j\right)^2 + I_i^2 I_l\left(1+\sum_{j=1}^{i-1} I_j\right) + I_i I_l\left(1+\sum_{j=1}^{i-1} I_j\right)\left(\sum_{r=i+1}^{l-1} I_r\right)\right)\\
&= \frac{p^2}{i \cdot l}\Big(2 + (3i-5+l)p + (i-1)(l-3)p^2\Big).
\end{aligned}$$

Therefore,

$$\begin{aligned}
c_{il} &= \frac{p^2}{i \cdot l}\Big(2 + (3i-5+l)p + (i-1)(l-3)p^2 - \Big(1+(i-1)p\Big)\Big(1+(l-1)p\Big)\Big) =\\
&= \frac{1}{l}p^2(1-p)\Big((1-2p)\frac{1}{i} + 2p\Big).
\end{aligned}$$

Summing up all the coefficients $\{c_{il}\}_{i,l=1}^{k}$ we obtain

$$\begin{aligned}
\sum_{i=1}^{k}\sum_{l=1}^{k} c_{il} &= \sum_{i=1}^{k} c_{ii} + 2\sum_{i=1}^{k-1}\sum_{l=i+1}^{k} c_{il} =\\
&= \sum_{i=1}^{k}\left(p(1-p)\Big(p^2 + p(3-2p)\frac{1}{i} + (1-3p+p^2)\frac{1}{i^2}\Big)\right) +\\
&\quad + 2\sum_{i=1}^{k-1}\sum_{l=i+1}^{k} p^2(1-p)\left((1-2p)\frac{1}{i \cdot l} + 2p\frac{1}{l}\right) =\\
&= k p^3(1-p) + p^2(3-2p)\sum_{i=1}^{k}\frac{1}{i} + p(1-3p+p^2)\sum_{i=1}^{k}\frac{1}{i^2} +\\
&\quad + 2p^2(1-p)(1-2p)\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\frac{1}{i \cdot l} + 4p^3(1-p)\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\frac{1}{l}.
\end{aligned}$$

Using the results from Lemma 1 we, finally, get

$$\begin{aligned}
\mathrm{Var}(AP@k) &= \frac{1}{k^2}\left(\sum_{i=1}^{k} c_{ii} + 2\sum_{i=1}^{k-1}\sum_{l=i+1}^{k} c_{il}\right) =\\
&= \frac{1}{k^2}\left(p(1-p)\sum_{i=1}^{k}\Big(p^2 + p(3-2p)\frac{1}{i} + (1-3p+p^2)\frac{1}{i^2}\Big) +\right.\\
&\quad \left. + 2p^2(1-p)\sum_{i=1}^{k-1}\sum_{l=i+1}^{k}\Big((1-2p)\frac{1}{i \cdot l} + 2p\frac{1}{l}\Big)\right) =\\
&= \frac{5}{k}p^3(1-p) + \frac{1}{k^2}p(1-p)\left(p(1-2p)\big(3H_k + H_k^2\big) + (1-p)(1-3p)H_k^{(2)}\right).
\end{aligned}$$

□

Acknowledgments

The authors thank anonymous referees for their careful reading of the first version of the manuscript and valuable remarks that helped to significantly improve the paper.

Funding

The second author, Tetiana Ianevych, was supported by the Ministry of Education and Science of Ukraine within the grant for the prospective development of the scientific field “Mathematical Sciences and Natural Sciences” at Taras Shevchenko National University of Kyiv.

References

[1] Agarwal, D.K., Chen, B.-C.: Evaluation Methods, pp. 55–78. Cambridge University Press (2016)

[2] Bestgen, Y.: Exact expected average precision of the random baseline for system evaluation. Prague Bull. Math. Linguist. **103**, 131–138 (2015). https://doi.org/10.1515/pralin-2015-0007

[3] Boas, R.P., Wrench, J.W.: Partial sums of the harmonic series. Am. Math. Mon. **78**(8), 864–870 (1971). MR0289994. https://doi.org/10.1080/00029890.1971.11992881

[4] Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: Proceedings of the Seventh International Conference on World Wide Web, pp. 107–117. Elsevier Science Publishers B. V. (1998). https://snap.stanford.edu/class/cs224w-readings/Brin98Anatomy.pdf

[5] Engström, C., Silvestrov, S.: Pagerank for networks, graphs and Markov chains. Theory Probab. Math. Stat. **96**, 61–83 (2017) MR3666872. https://doi.org/10.1090/tpms/1034

[6] Gebremeskel, G.G., de Vries, A.P.: Recommender systems evaluations: offline, online, time and a/a test. In: Balog K., F.N.M.C. Cappellato L. (ed.) Working Notes of CLEF 2016 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings vol. 1609, pp. 642–656 (2016). https://ceur-ws.org/Vol-1609/16090642.pdf

[7] Jadon, A., Patil, A.: A Comprehensive Survey of Evaluation Techniques for Recommendation Systems (2024). https://arxiv.org/abs/2312.16015

[8] Knuth, D.E.: Harmonic Numbers, 1st edn., pp. 73–78. Addison-Wesley (1968)

[9] Valcarce, D., Bellogín, A., Parapar, J., Castells, P.: Assessing ranking metrics in top-n recommendation. Inf. Retr. J. **23**(4), 411–448 (2020). https://doi.org/10.1007/s10791-020-09377-x