Additional Data for Two-Step Shortlisting by Imperfect Experts

Stephan Kolassa
kolassa@minet.uni-jena.de
Institut für Angewandte Mathematik
Friedrich-Schiller-Universität Jena
Germany

April 5, 2004

Abstract

We give additional data from Monte Carlo simulations of the two-step shortlisting model with imperfect experts proposed in our previous report. We consider uniformly, normally and exponentially distributed true values and imprecisions and give empirical hitting ratios as well as empirical expected results for different imprecisions.

MSC 2000: Primary 91B06, Secondary 90B50

1 Introduction

Decision-making processes often consist of more than one step, for instance:

- Junior management or consultants analyze a large number of possible approaches to a problem and present a short list of alternatives. Senior management then reviews this list and makes a final choice or decision.
- Computerized Decision Support Systems such as *Recommender Systems* (Resnick and Varian, 1997) make a small set of recommendations from which the end user selects one course of action.
- Chess professionals use the k-best mode of computer chess programs and make the final choice themselves when analyzing games or researching new variants (see Bareev, 2001).
- Since computer Go programs are still very weak compared to human players, there have been experiments in which one or more programs proposed the $k \approx 50$ moves with the highest internal valuation, from which a human player selected one move to actually play (Althöfer and Kolassa, 2003).

- Among all applicants to a job, usually only a small number is invited to an interview, based on their written applications and curricula vitae.

  This process of selection is called *shortlisting*. We adopt this term in this report.

These examples are instances of "division of labor"-approaches for making decisions based on two-step algorithms.

However, real world experts are not perfect. They cannot measure hidden costs and benefits of a course of action and need to reduce complexity to be able to handle reality. One implicit assumption of two-step procedures is that the sequential application of more than one expert improves the final decision. This assumption is rarely explicitly acknowledged.

2 Overview

Kolassa and Schwarz (2003) proposed the following model of two-step shortlisting by imperfect experts:

Model 1. Consider $3n$ independently distributed random variables

$$\begin{aligned} x_1, \ldots, x_n & \text{ uniformly distributed in } [0,1], \\ \gamma_1, \ldots, \gamma_n & \text{ uniformly distributed in } [0,\Gamma], \\ \text{and } \delta_1, \ldots, \delta_n & \text{ uniformly distributed in } [0,\Delta]. \end{aligned}$$

We consider two experts, Alice and Bob. Alice observes the set $\{x_i + \gamma_i \,|\, 1 \leq i \leq n\}$ and constructs a shortlist $\mathcal{S} \subseteq \{1, \ldots n\}$ consisting of the indices of her k largest observations $x_i + \gamma_i$.

Bob now observes the set $\{x_i + \delta_i \,|\, i \in \mathcal{S}\}$ and selects the index $\tilde{\imath}$ of his largest observation.

The correct index i^* Alice and Bob are looking for is the index of the maximal x_i. ♢

We note that i^* is almost surely well-defined and that the zero set on which it is not will not influence our results.

The rationale of this model is that the x_i quantify some "real" value of the i^{th} alternative and that the γ_i and δ_i are imprecisions in Alice and Bob's observation of the data. The larger Γ or Δ are, the less precise are the associated experts. We will restrict ourselves to the case $\Gamma = \Delta$, i.e., Alice is as precise as Bob.

We are interested in the probability that this process returns the index i^* of the "real" maximum $\max_{1 \leq i \leq n} x_i$: the *hitting ratio*

$$h(n \xrightarrow{\Gamma} k \xrightarrow{\Gamma} 1).$$

Monte Carlo simulations yield the *empirical hitting ratio* as an estimation for the hitting ratio:

$$h^{\text{emp}}(n \xrightarrow{\Gamma} k \xrightarrow{\Gamma} 1) := \frac{\#\{\text{Monte Carlo runs in which } i^* \text{ is returned}\}}{\#\{\text{all Monte Carlo runs}\}}.$$

In addition, we also consider the expected value of the true value returned $\mathbb{E}(x_{\tilde{\imath}})$, the *expected result*. The *empirical expected result* is calculated as the empirical hitting ratio based on Monte Carlo simulations.

Although a theoretical analysis of Model 1 is rather complicated, Kolassa and Schwarz (2003) as well as Kolassa (2004) already reported results of Monte Carlo simulations with varying values of n and Γ. They also considered non-uniformly distributed true values and imprecisions.

In this paper, we report on further Monte Carlo experiments. We consider uniformly, normally and exponentially distributed true values and imprecisions with different values of the imprecision parameter Γ: specifically, we examine $\Gamma \in \{0.01, 0.1, 1, 10, 100\}$. We always consider $n = 100$ alternatives as a compromise between keeping computation time low and having "enough room" for interesting phenomena to appear.

For every value of Γ, $R = 10^6$ Monte Carlo simulations were conducted. In each run, new sets of variables x_i, γ_i and δ_i were generated, and the shortlisting was conducted as detailed above.

We denote a uniformly distributed random variable x in the interval $[a, b]$ by $x \sim U[a, b]$ and generate non-uniform distributions in the following ways:

- To generate normally distributed random variables, we apply the ratio method as discovered by Kinderman and Monahan (1976, 1977) – see also Knuth (1998, Section 3.4.1). We denote a normally distributed random variable x with mean μ and standard deviation σ by $x \sim N(\mu, \sigma^2)$.

- To generate exponentially distributed random variables, we directly use the logarithm method (see Knuth, 1998, Section 3.4.1): if $u \sim U[0, 1]$, then $x := -\beta \log u$ is exponentially distributed with mean β. We denote an exponentially distributed random variable x with scale parameter β by $x \sim E(\beta)$, i. e., x has the probability density

$$f(z) = \begin{cases} \frac{1}{\beta} e^{-\frac{z}{\beta}}, & \text{if } z \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

  Then x has mean $\mathbb{E}(x) = \beta$ and standard deviation $\sigma_x = \beta$.

However, the regularity problems of many random number generators are well known – the findings of Park and Miller (1988) are unfortunately still revelant. We therefore conducted all simulations twice, using two different well accepted random number generators: a generator based on a combination of four linear congruence generators as proposed by L'Ecuyer and Andres (1997) and a generator as described by Press et al. (1992, p. 282). The results were qualitatively the same, and we will only give the results of the simulations that used L'Ecuyer and Andres' generator.

All computations were performed on a Compaq AlphaServer DS20E with two Alpha EV6.7 (21264A) processors at 667 MHz and 4 GB RAM running Compaq Tru64 UNIX V5.1 or on a PC with an AMD Athlon XP 2800+ processor and 512 MB RAM running SuSE Linux 8.0.

γ_i, δ_i \ x_i	$U[0,1]$		$N(0,1)$		$E(1)$	
	HR	ER	HR	ER	HR	ER
$U[0,\Gamma]$	Figure 2	Figure 3	Figure 8	Figure 9	Figure 14	Figure 15
$N(0,\Gamma^2)$	Figure 4	Figure 5	Figure 10	Figure 11	Figure 16	Figure 17
$E(\Gamma)$	Figure 6	Figure 7	Figure 12	Figure 13	Figure 18	Figure 19

Table 1: Which Figure shows which result? We abbreviate "HR" for the empirical hitting ratio and "ER" for the empirical expected result.

3 Results and Discussion

In this section, we comment on the results in some detail.

To make nomenclature easier, we abbreviated the distributions: "U" stands for the uniform distribution, "N" for the normal distribution, and "E" for the exponential distribution. For instance, "the case (U, E)" refers to $x_i \sim U[0,1]$ and $\gamma_i, \delta_i \sim E(\Gamma)$.

The results themselves are shown in graphical form at the end of this paper. Table 1 summarizes which results are shown in which Figure, and Table 2 on page 7 provides a quick overview of the results.

Upon visual inspection of the results, three "geographical features" stand out:

Peaks: Peaks are sharply defined interior maxima of the empirical hitting ratio or the empirical expected result. In the case of a peak, the shortlisting process improves the output, and the best value of the shortlist size k is uniquely defined.

Valleys: Valleys are interior minima of the empirical hitting ratio or the empirical expected result. At a valley, the shortlisting yields worse outputs than at neighboring values of k.

Tables: Tables are extended plateaus of the empirical hitting ratio or the empirical expected result. The shortlisting process improves the output, but the optimal value of k is not clear.

The motivation for the term "Table" is the geographical formation known as "table mountains" or "mesas," cf. Figure 1. Another possible name for this phenomenon would have been "plateau", but this word cannot be distinguished from a Peak by its first letter alone.

In the case of Peaks and Tables, an appealing hypothesis would be that the hitting ratio is unimodal in k, i. e., it first rises and then falls. However, there are also values of Γ and n for which both Valleys and Peaks appear, cf. Kolassa and Schwarz (2003, Figures 1 and 2).

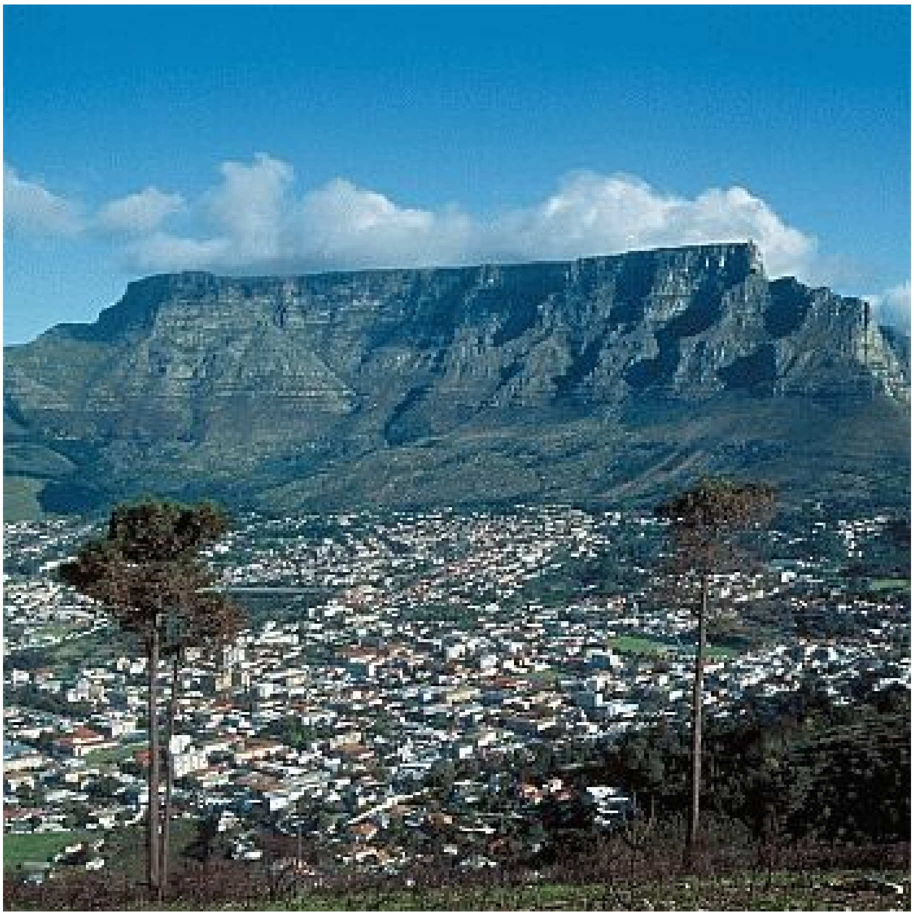

Figure 1: The table mountain near Kapstadt, South Africa. Source: SKR – Studien Kontakt Reisen, http://www.skr.de/themen/suedafrika/suedafrika-gesichter/

Let us consider the results in detail:

The case (U, U), cf. Figures 2 and 3: The empirical hitting ratio has a Peak for $\Gamma = 0.01$ and $\Gamma = 0.1$. A Valley appears for $\Gamma = 1$, $\Gamma = 10$ and $\Gamma = 100$. The empirical hitting ratio falls quite rapidly with growing Γ. For $\Gamma = 10$, the empirical hitting ratio nearly halves from 3.9 % for $k = 1$ to 2.0 % for $k = 10$.

The empirical expected result has a Peak for $\Gamma = 0.1$ and a Valley for $\Gamma = 1$, $\Gamma = 10$ and $\Gamma = 100$.

The case (U, N), cf. Figures 4 and 5: The empirical hitting ratio has a Peak for $\Gamma = 0.01$, $\Gamma = 0.1$, $\Gamma = 1$ and (with a high level of noise) for $\Gamma = 10$. The empirical hitting ratio again falls quickly with increasing Γ.

The empirical expected result has a Peak for all values of Γ considered, although the noise level increases for $\Gamma = 100$.

The case (U, E), cf. Figures 6 and 7: The empirical hitting ratio shows a Peak for $\Gamma = 0.01$ and $\Gamma = 0.1$. For $\Gamma = 1$, a Table appears, and it is blurred but still visible for $\Gamma = 10$.

The empirical expected result has a Peak for $\Gamma = 0.01$. Starting with $\Gamma = 0.1$, a Table appears and persists for $\Gamma = 1$ and $\Gamma = 10$. The highest difference between the minimum and the maximum empirical expected result occurs for the table formation for $\Gamma = 1$.

The case (N, U), cf. Figures 8 and 9: In the empirical hitting ratio, a Peak is blurred for $\Gamma = 0.1$ and clearly visible for $\Gamma = 1$. For $\Gamma = 10$ and $\Gamma = 100$, a Valley appears. This Valley is quite pronounced: the hitting ratio decreases from 15.1 % for $k = 1$ to 8.9 % for $k = 9$ for $\Gamma = 10$ and from 3.2 % for $k = 1$ to 1.5 % for $k = 10$ for $\Gamma = 100$.

The empirical expected result has a Peak for $\Gamma = 1$ and a Valley for $\Gamma = 10$ and $\Gamma = 100$. The Valley again is rather deep: for $\Gamma = 100$, the empirical expected result decreases from 0.64 for $k = 1$ to 0.19 for $k = 11$.

The case (N, N), cf. Figures 10 and 11: Peaks are visible in the empirical hitting ratio for $\Gamma = 0.1$, $\Gamma = 1$ and $\Gamma = 10$.

The empirical expected result shows Peaks for $\Gamma = 0.1$ (blurred), $\Gamma = 1$, $\Gamma = 10$ and $\Gamma = 100$ (blurred).

The case (N, E), cf. Figures 12 and 13: The empirical hitting ratio shows a Peak for $\Gamma = 0.1$ and $\Gamma = 1$. For $\Gamma = 10$, a Table appears.

For the empirical expected result, Peaks are visible for $\Gamma = 0.1$ and $\Gamma = 1$. A Table can be seen for $\Gamma = 10$ and (with more noise) for $\Gamma = 100$.

The case (E, U), cf. Figures 14 and 15: For $\Gamma = 1$, the empirical hitting ratio has a Peak. For $\Gamma = 10$ and $\Gamma = 100$, a Valley appears.

In the empirical expected result, a Peak can be guessed at for $\Gamma = 1$. Valleys are obvious for $\Gamma = 10$ and $\Gamma = 100$.

The case (E, N), cf. Figures 16 and 17: The empirical hitting ratio has a Peak for $\Gamma = 0.1$, $\Gamma = 1$, $\Gamma = 10$ and (indistinctly) for $\Gamma = 100$.

The empirical expected result shows a Peak for $\Gamma = 1$ and $\Gamma = 10$. For $\Gamma = 100$, a slightly blurred Peak appears.

The case (E, E), cf. Figures 18 and 19: A Peak appears in the empirical hitting ratio for $\Gamma = 0.1$ and $\Gamma = 1$. For $\Gamma = 10$ and $\Gamma = 100$, a Table can be seen. For $\Gamma = 100$, the Table is blurred by noise. For $\Gamma = 1$, the Peak is quite high: the empirical hitting ratio increases from 35 % for $k = 1$ to 62 % for $k = 5$.

The empirical expected result has a peak for $\Gamma = 1$ and a Table for $\Gamma = 10$. For $\Gamma = 100$, a large Table appears for $2 \leq k \leq 98$, in spite of a high level of noise.

The above results are summarized in Table 2.

γ_i, δ_i \ x_i		$U[0,1]$		$N(0,1)$		$E(1)$	
distribution	Γ	HR	ER	HR	ER	HR	ER
$U[0,\Gamma]$	0.01	P	?	?	?	?	?
	0.1	P	P	P?	?	?	?
	1	V	V	P	P	P	P?
	10	V	V	V	V	V	V
	100	V	V	V	V	V	V
$N(0,\Gamma^2)$	0.01	P	P	?	?	?	?
	0.1	P	P	P	P?	P	?
	1	P	P	P	P	P	P
	10	P?	P	P	P	P	P
	100	?	P?	?	P?	P?	P?
$E(\Gamma)$	0.01	P	P	?	?	?	?
	0.1	P	T	P	P	P	?
	1	T	T	P	P	P	P
	10	T?	T	T	T	T	T
	100	?	?	?	T?	T?	T?

Table 2: Where does which kind of "geographical" formation appear? "P" stands for a Peak, "V" for a Valley, "T" for a Table and "?" for a high level of noise.

Comparing the results for the different distributions and the different values of Γ as given in Table 2, a number of general conclusions can be drawn.

- There is no qualitative difference between the "geography" of the empirical hitting ratio and of the empirical expected result. All geographical features appear either in both or in neither of the corresponding Figures, possibly for different values of Γ.
- Peaks are predominant. For all combinations of distributions, there are some values of Γ for which Peaks appear. Thus, two-step shortlisting is in a way vindicated: if the imprecisions are "good", shortlisting will improve the results.
- The geography of the results depends more on the distribution of the imprecisions than on the distribution of the true values:
  - Valleys appear only for uniformly distributed imprecisions.
  - Tables appear only for exponentially distributed imprecisions.

- Both very small and very large imprecisions have a tendency to produce a high level of noise. However, this again depends on the distribution of the imprecisions:
  - Uniformly distributed imprecisions tend to be noisy only for small Γ.
  - Normally and exponentially distributed imprecisions show noise for small and large Γ.

  For small Γ, there is less noise if the true values are uniformly distributed than if they are normally or exponentially distributed.

References

Althöfer, I. and S. Kolassa, 2003. Einer für alle, alle für einen. Das 52-Hirn auf dem Jenaer Turnier. Deutsche Go-Zeitung, Nr. 2/2003: 32–33.

Bareev, E., 2001. Electronic chess alchemy. ChessBase Magazine 82. URL http://www.chessbase.com/newsdetail.asp?newsid=1020.

Kinderman, A. J. and J. F. Monahan, 1976. Generating Random Variables from the Ratio of Two Uniform Deviates. In *Computer Science and Statistics: Proceedings of the Ninth Annual Symposium on the Interface*, pp. 197–199. Prindle, Weber and Schmidt, Boston.

Kinderman, A. J. and J. F. Monahan, 1977. Computer Generation of Random Variables Using the Ratio of Uniform Deviates. *ACM Transactions on Mathematical Software*, 3(3):257–260.

Knuth, D. E., 1998. *The Art of Computer Programming. Vol. 2: Seminumerical Algorithms*. Addison-Wesley, Reading, 3rd edition.

Kolassa, S., 2004. An Anomaly in Two-Step Shortlisting by Imperfect Experts. Submitted to *Mathematics and Computers in Simulation*.

Kolassa, S. and S. Schwarz, 2003. Two-Step Shortlisting by Imperfect Experts. *Technical Report 03-19*, Faculty of Mathematics and Computer Science, Friedrich-Schiller-Universität Jena, Germany. URL http://www.minet.uni-jena.de/preprints/kolassa_03/two_agent_shortlisting.pdf.

L'Ecuyer, P. and T. Andres, 1997. A Random Number Generator Based on the Combination of Four LCGs. *Mathematics and Computers in Simulation*, 44:99–107.

Park, S. K. and K. W. Miller, 1988. Random Number Generators: Good Ones are Hard to Find. *Communications of the ACM*, 31(10):1192–1201.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition.

Resnick, P. and H. R. Varian, 1997. Recommender Systems. *Communications of the ACM*, 40(3):56–58.

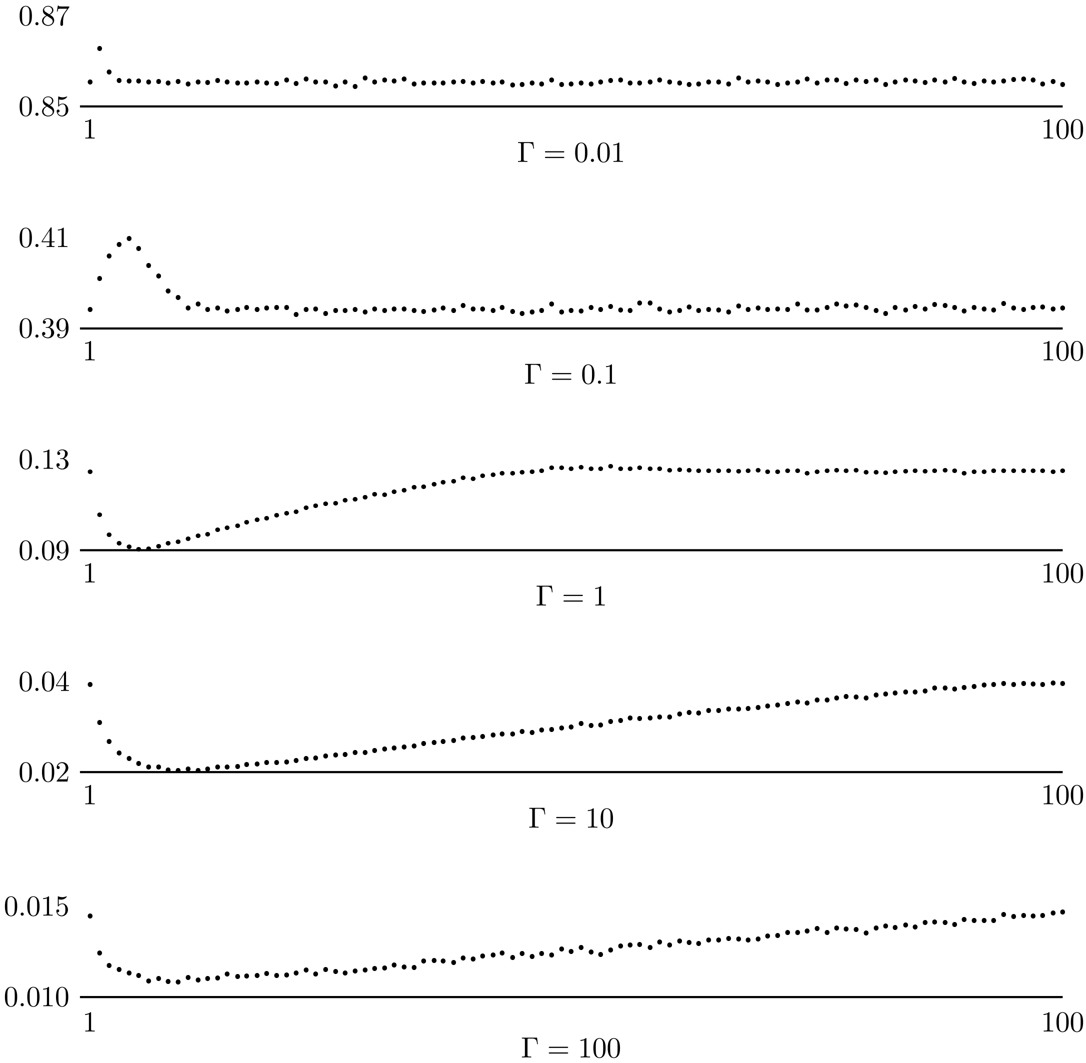

Figure 2: (U, U): the empirical hitting ratio

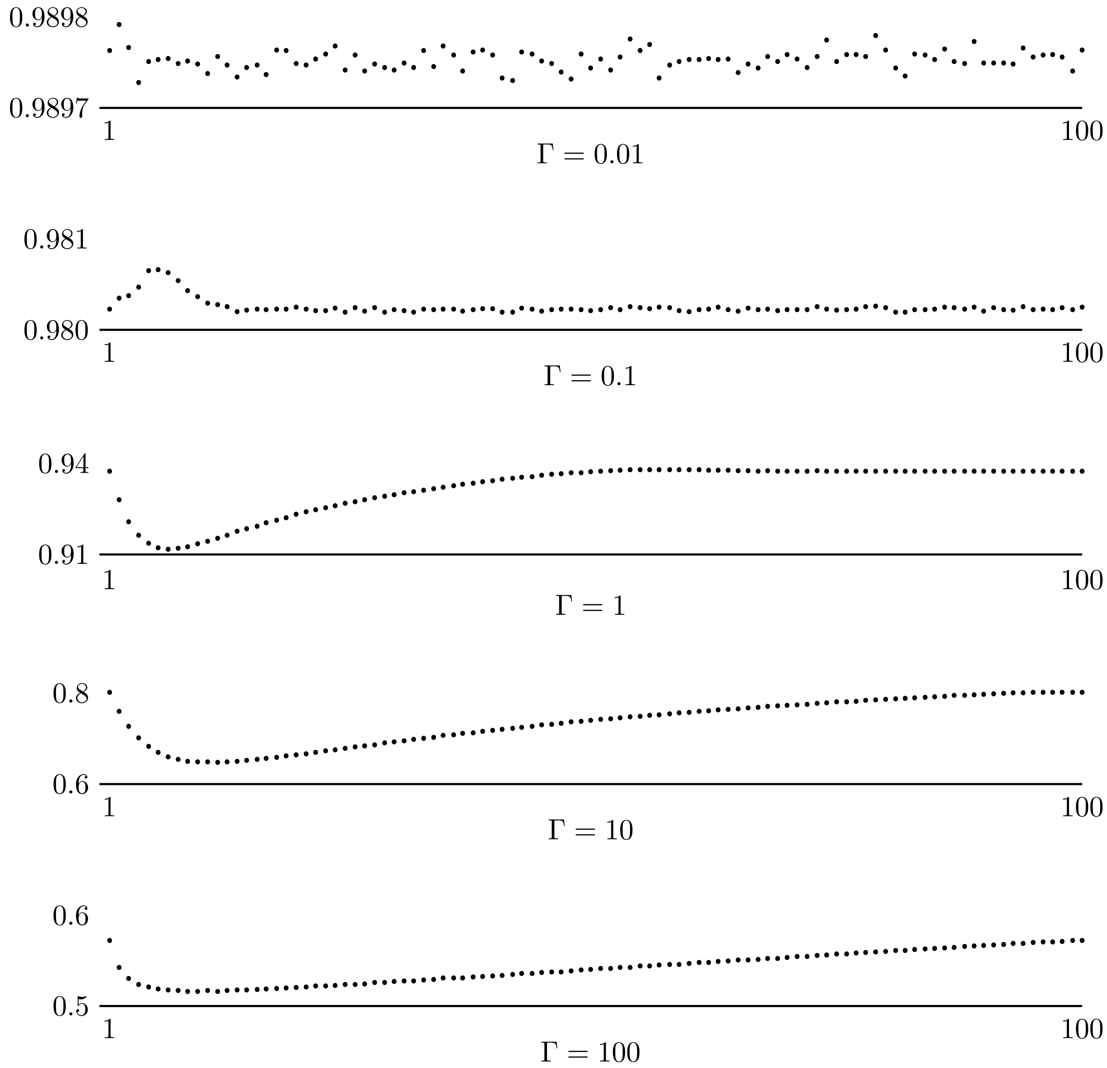

Figure 3: (U, U): the empirical expected result

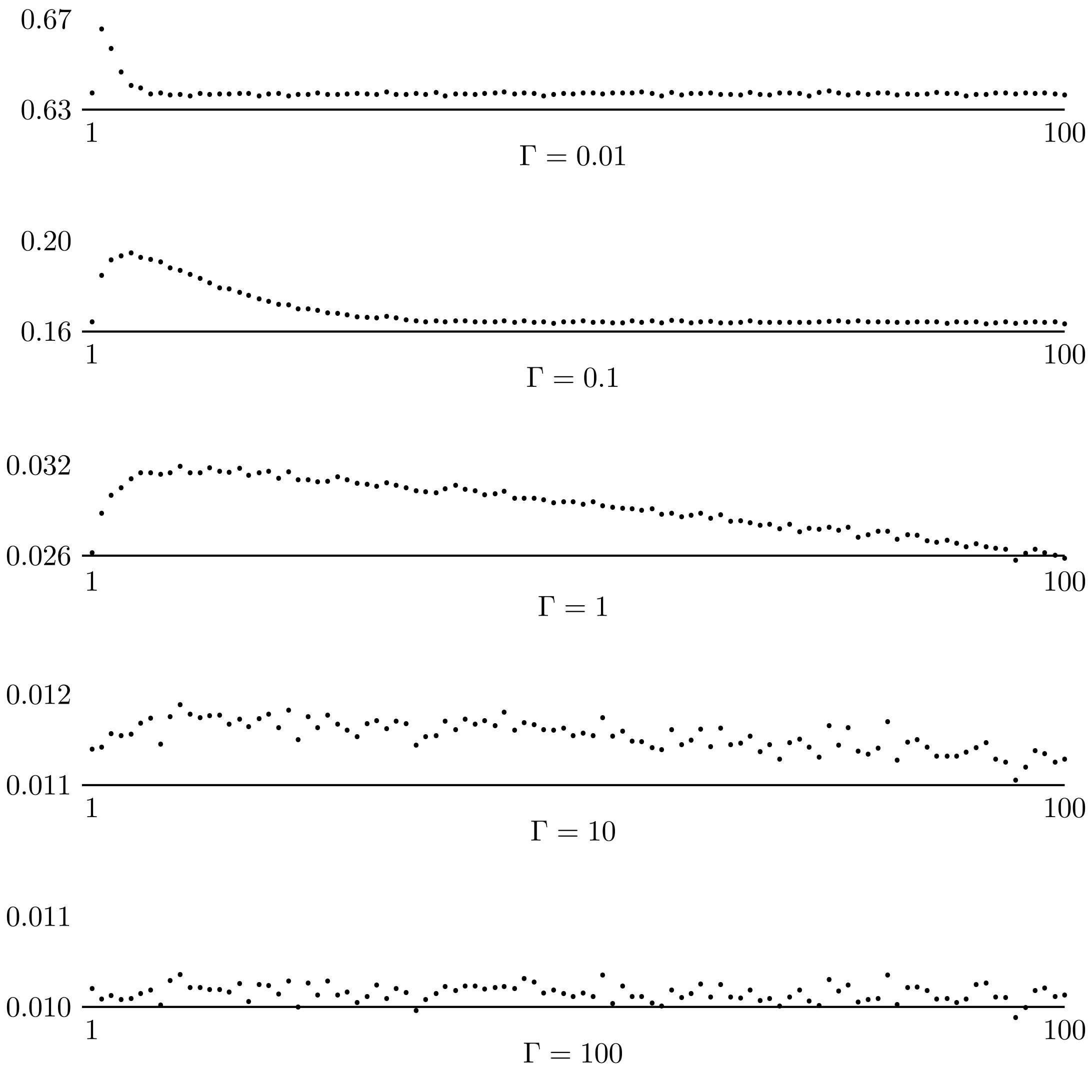

Figure 4: (U, N): the empirical hitting ratio

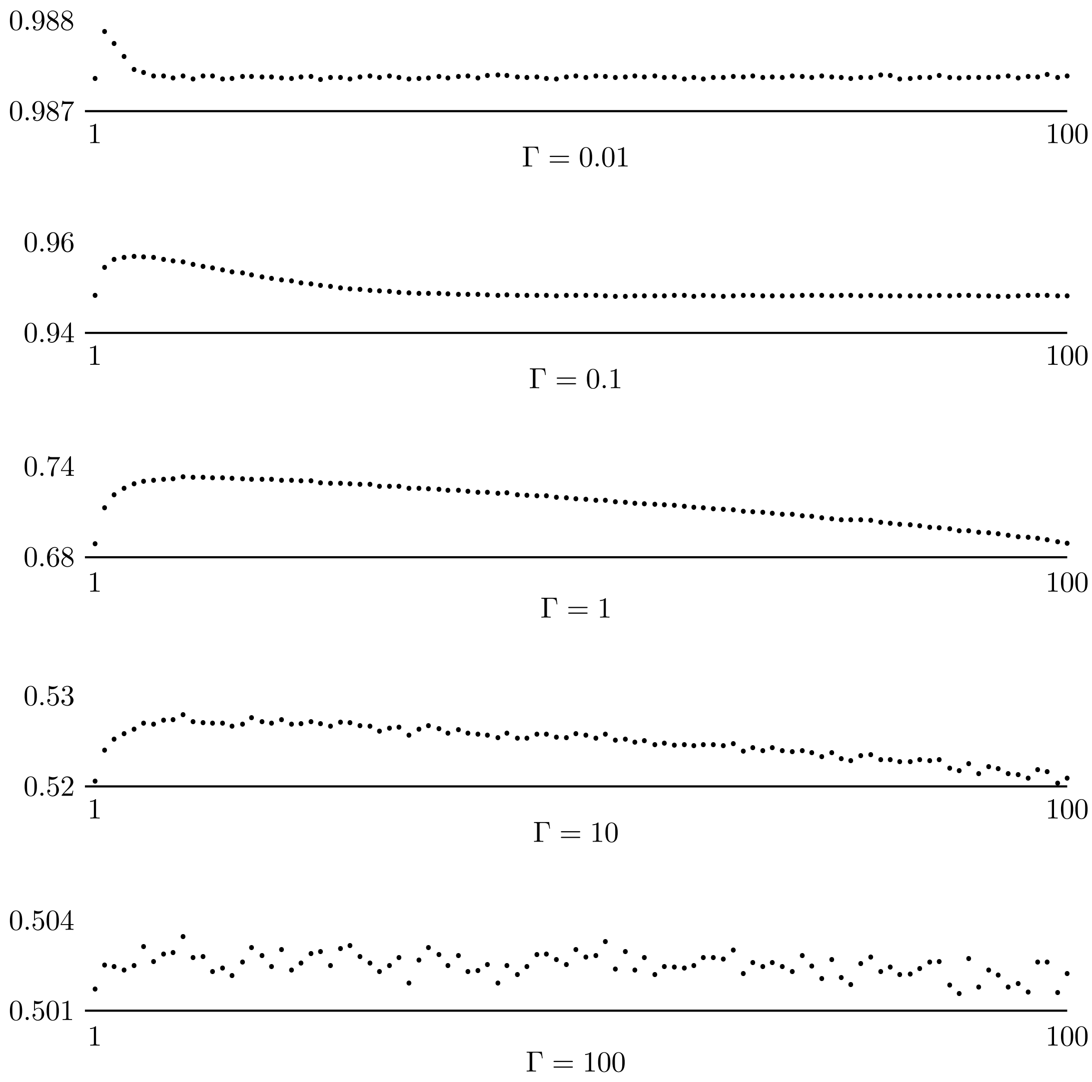

Figure 5: (U, N): the empirical expected result

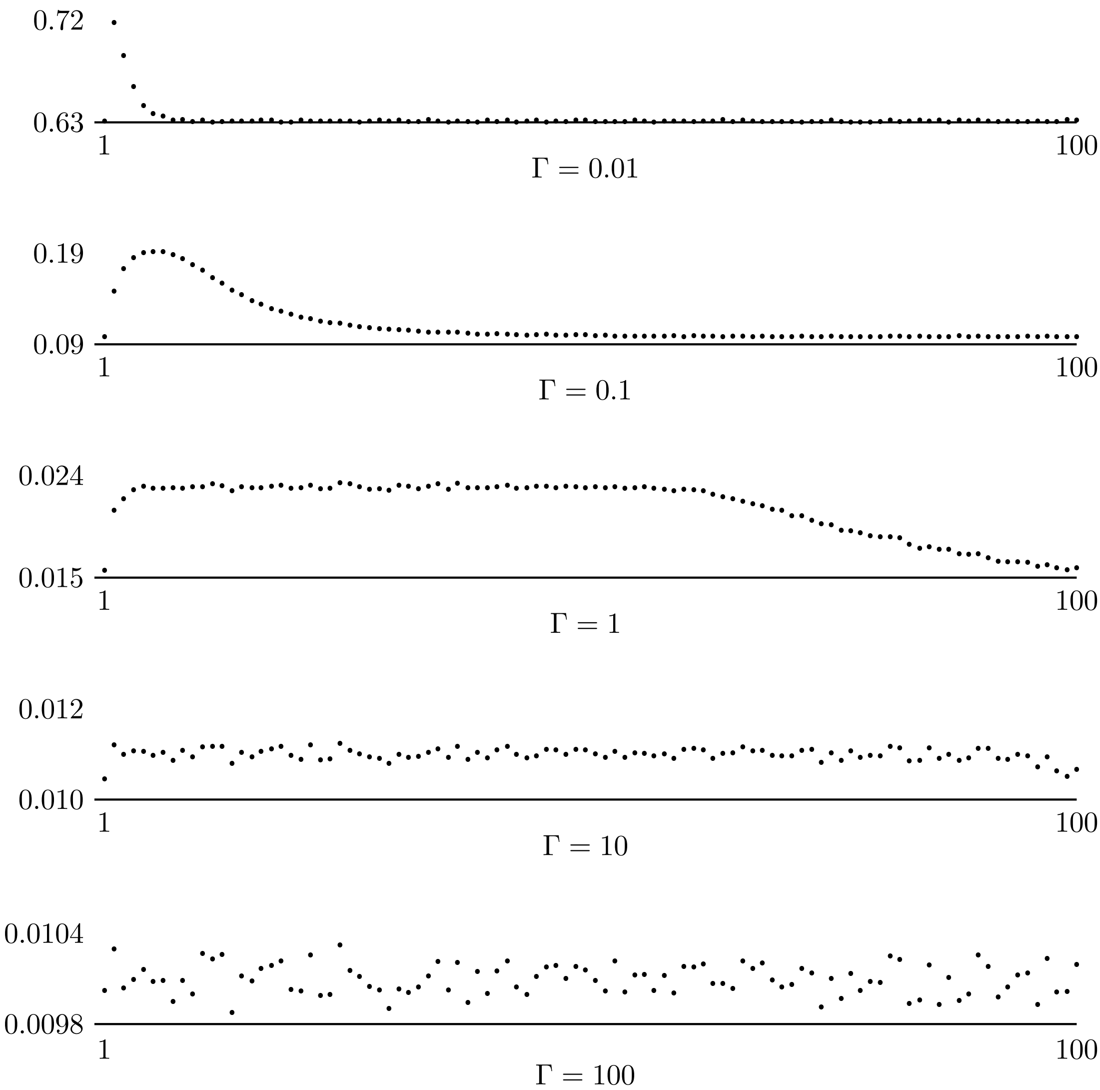

Figure 6: (U, E): the empirical hitting ratio

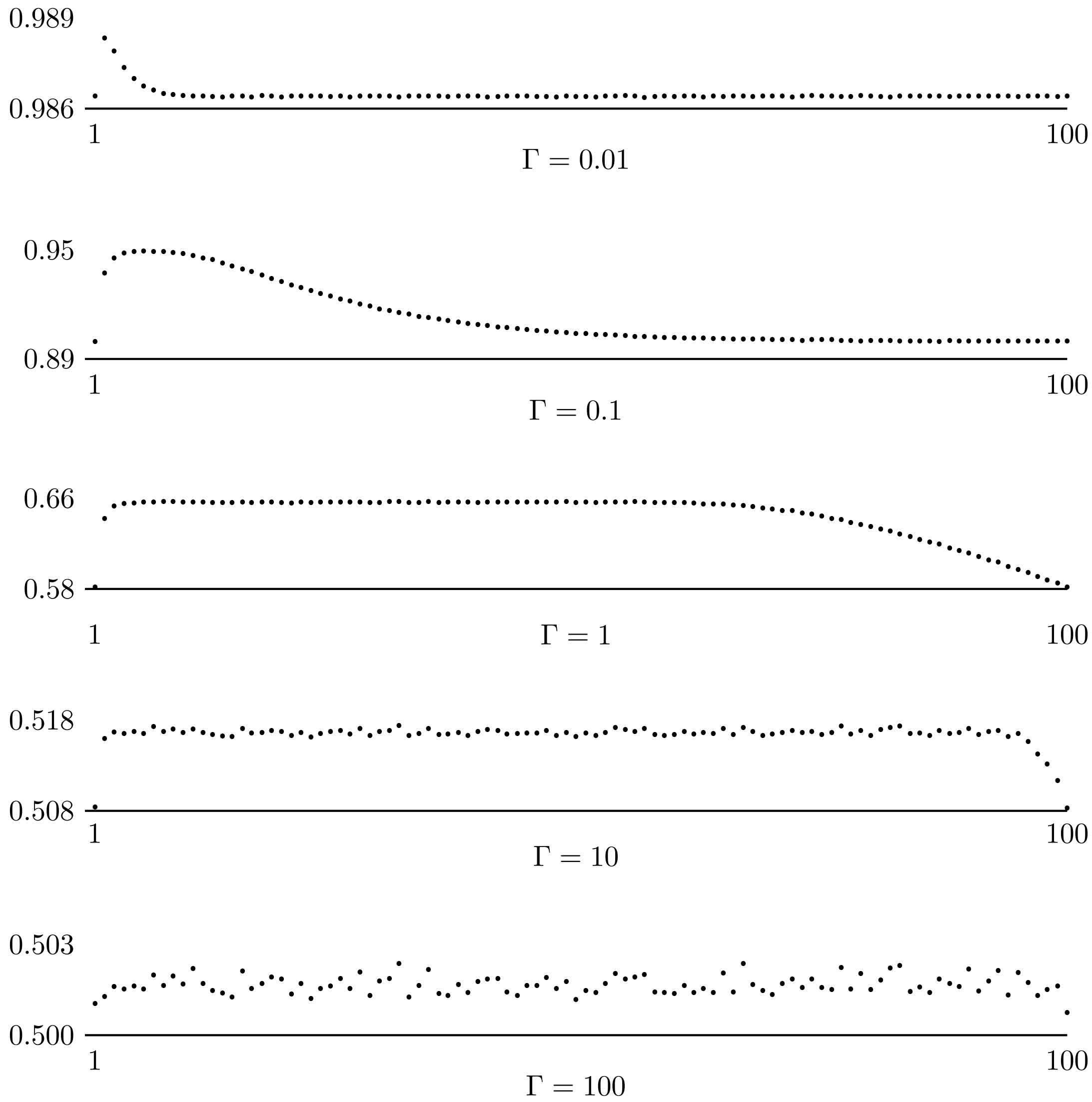

Figure 7: (U, E): the empirical expected result

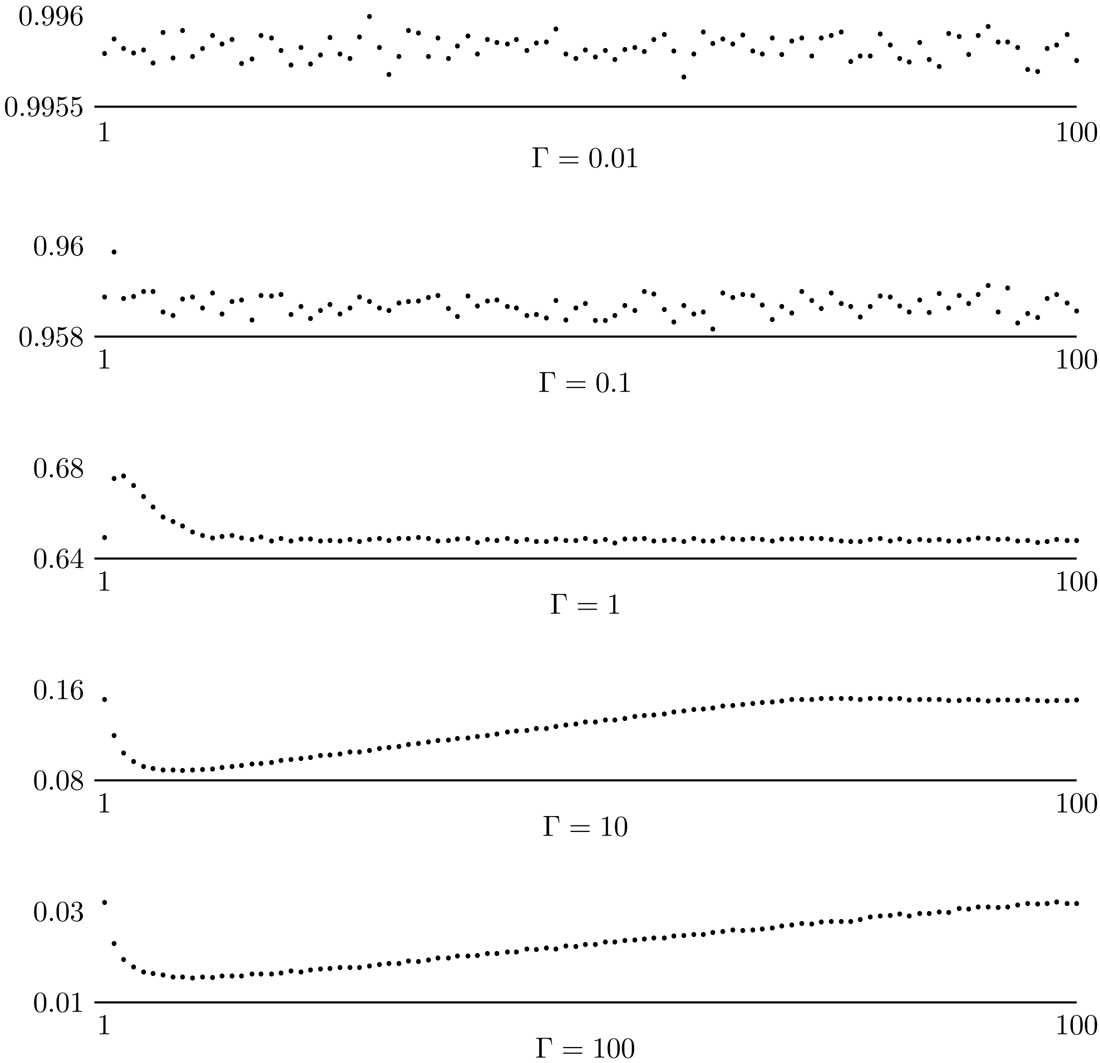

Figure 8: (N, U): the empirical hitting ratio

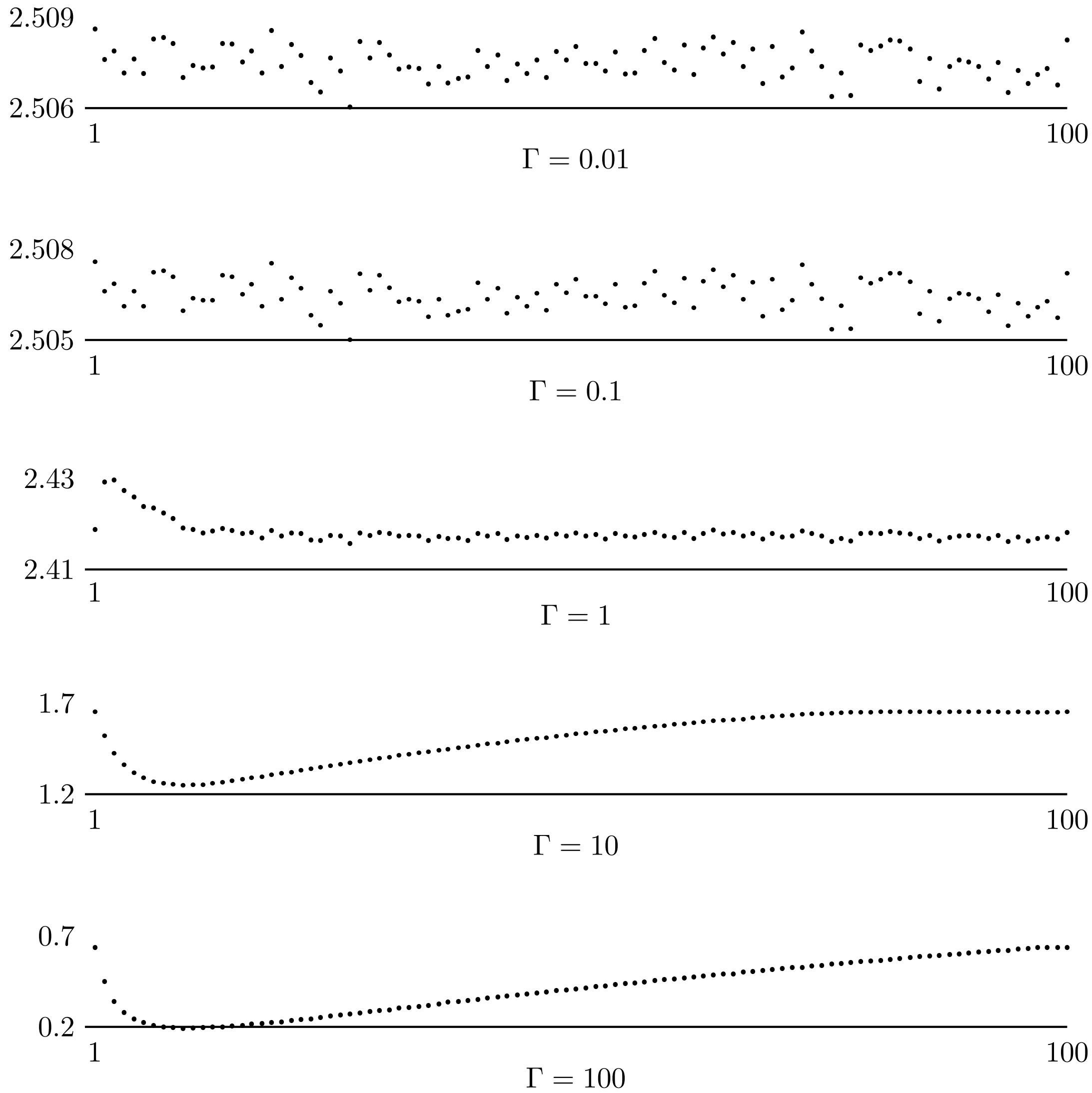

Figure 9: (N, U): the empirical expected result

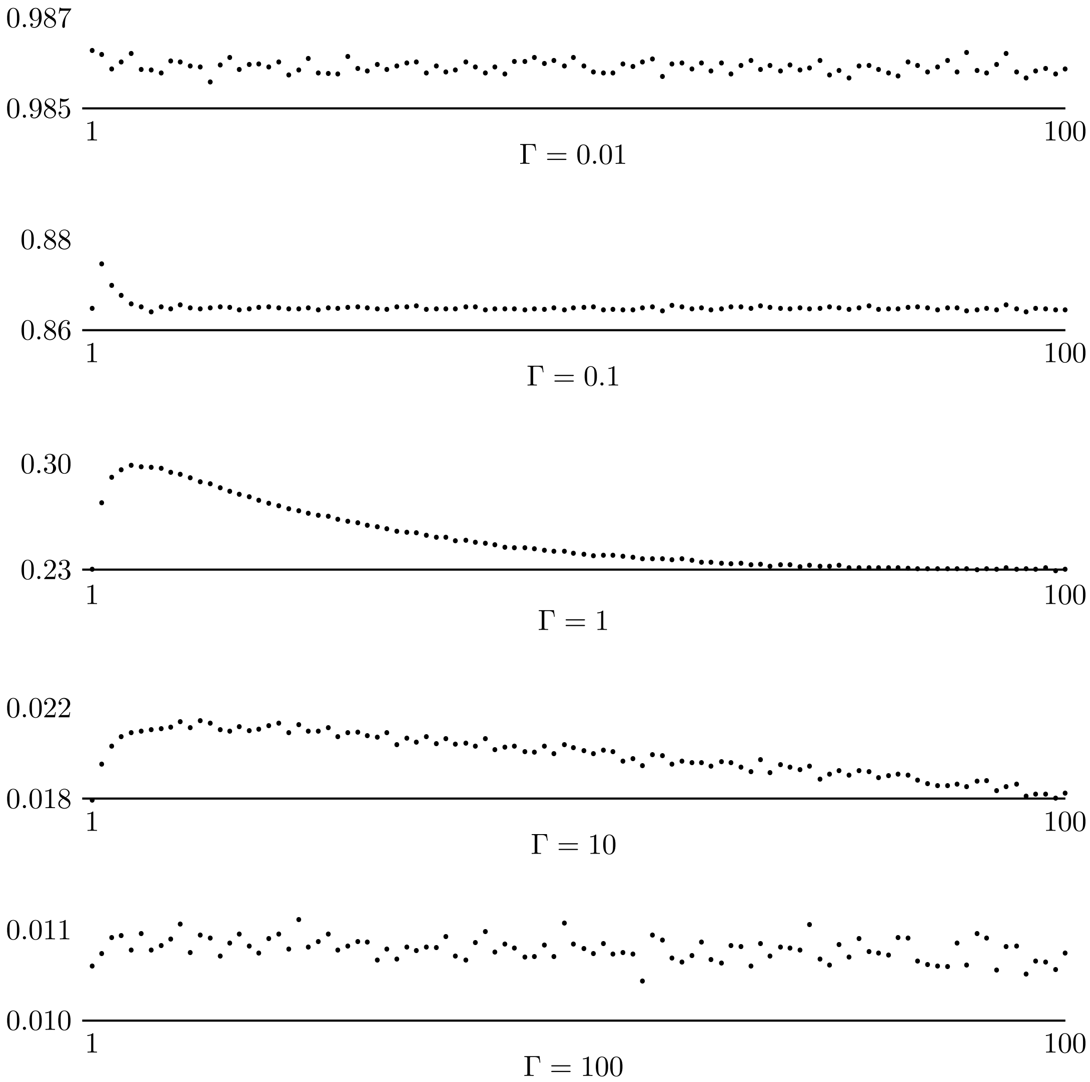

Figure 10: (N, N): the empirical hitting ratio

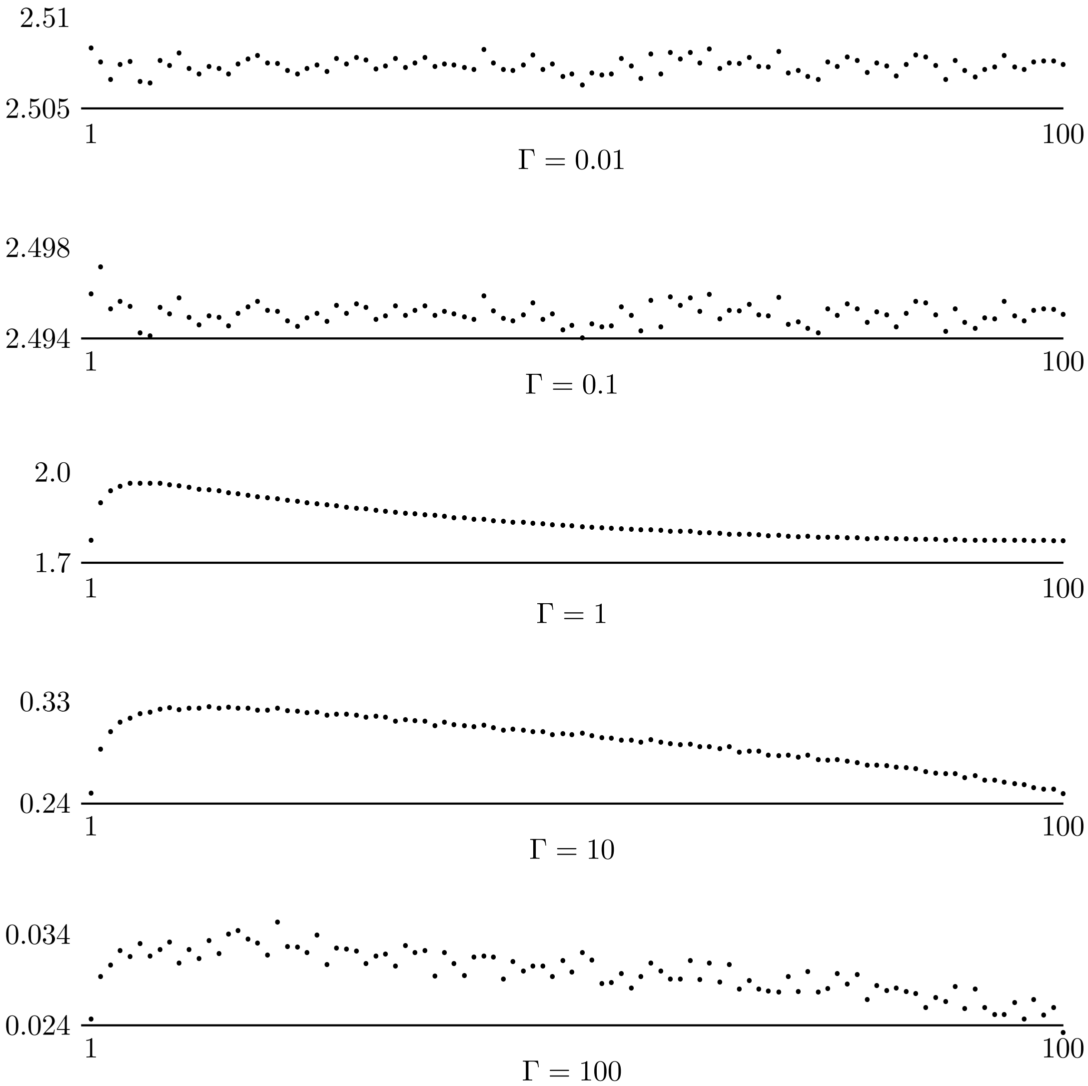

Figure 11: (N, N): the empirical expected result

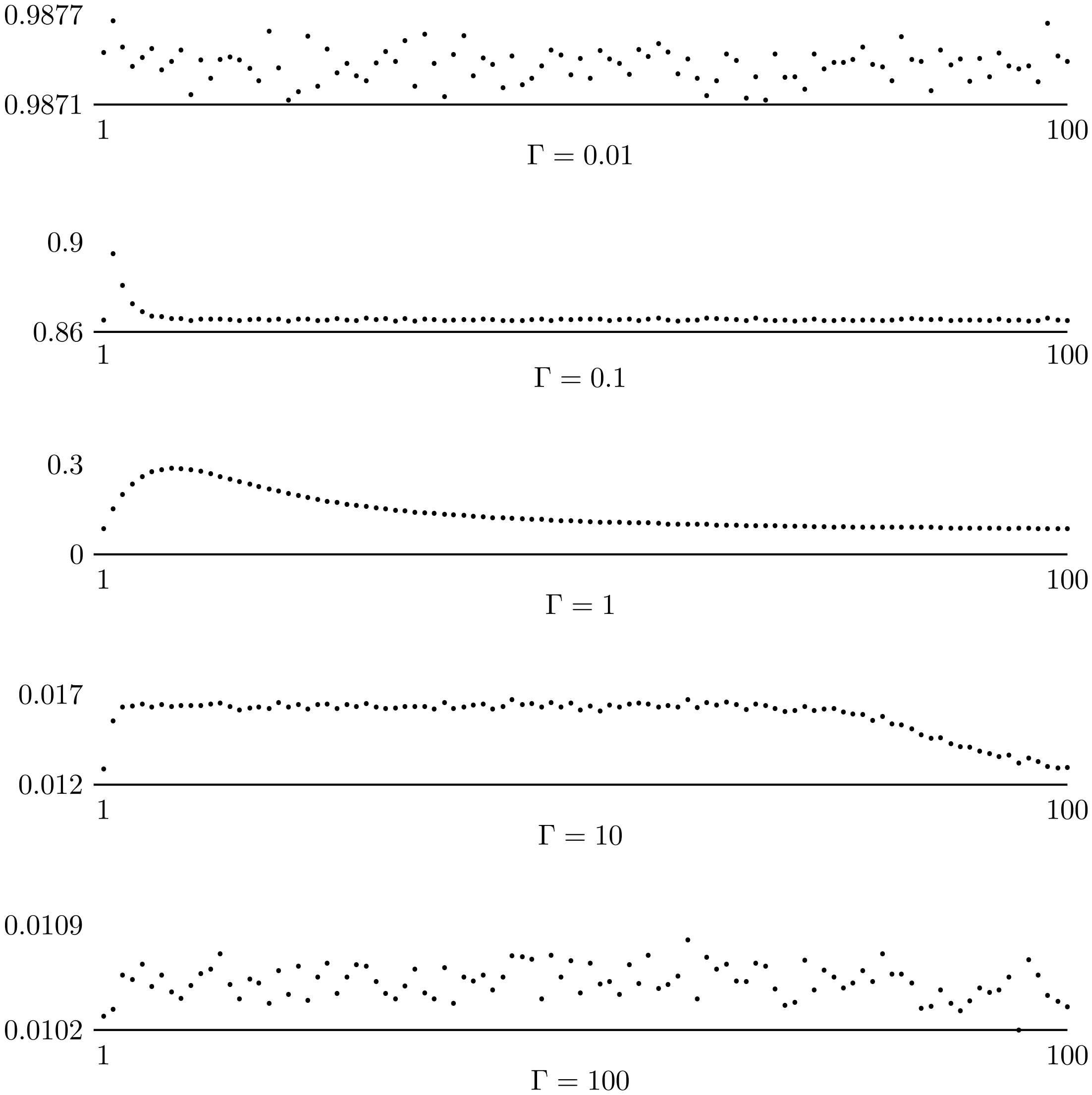

Figure 12: (N, E): the empirical hitting ratio

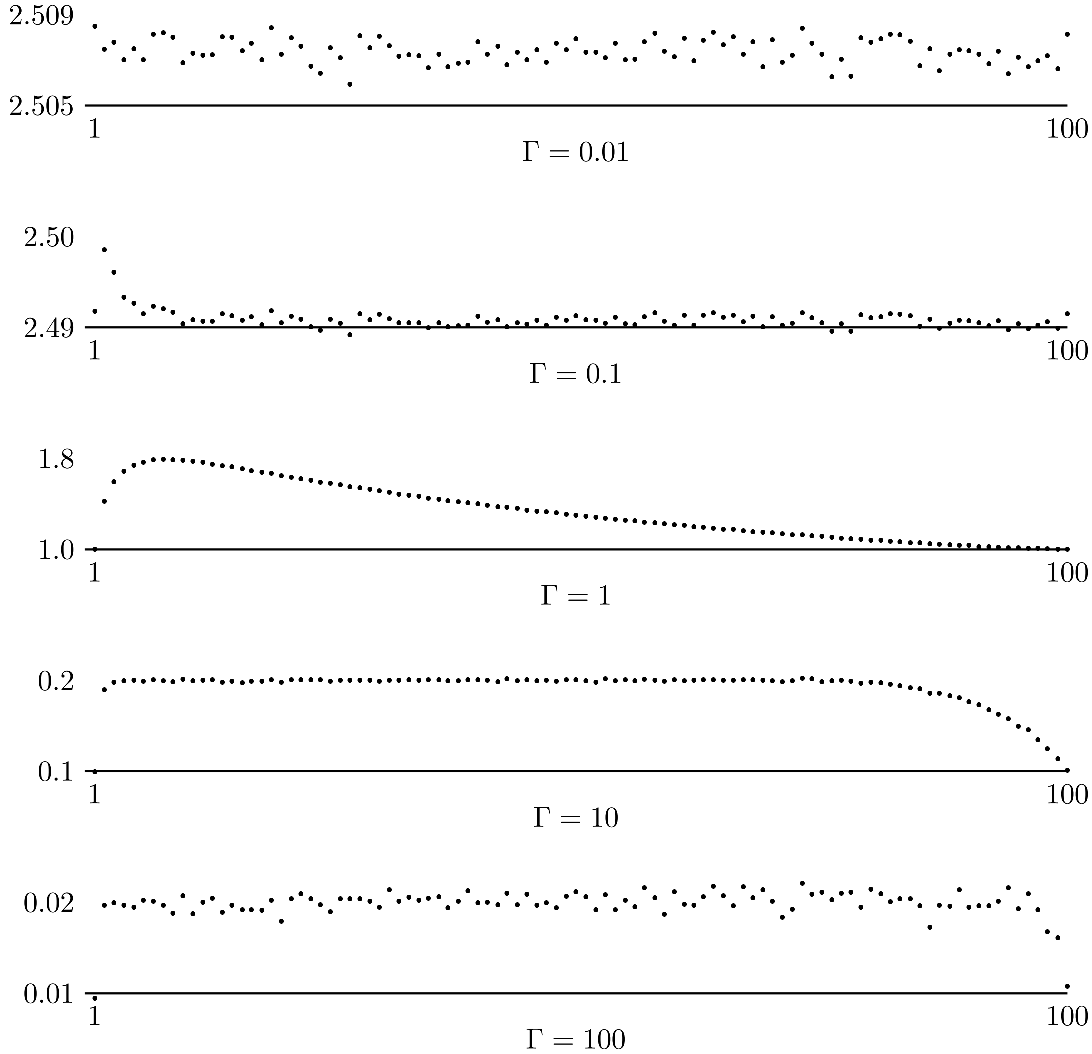

Figure 13: (N, E): the empirical expected result

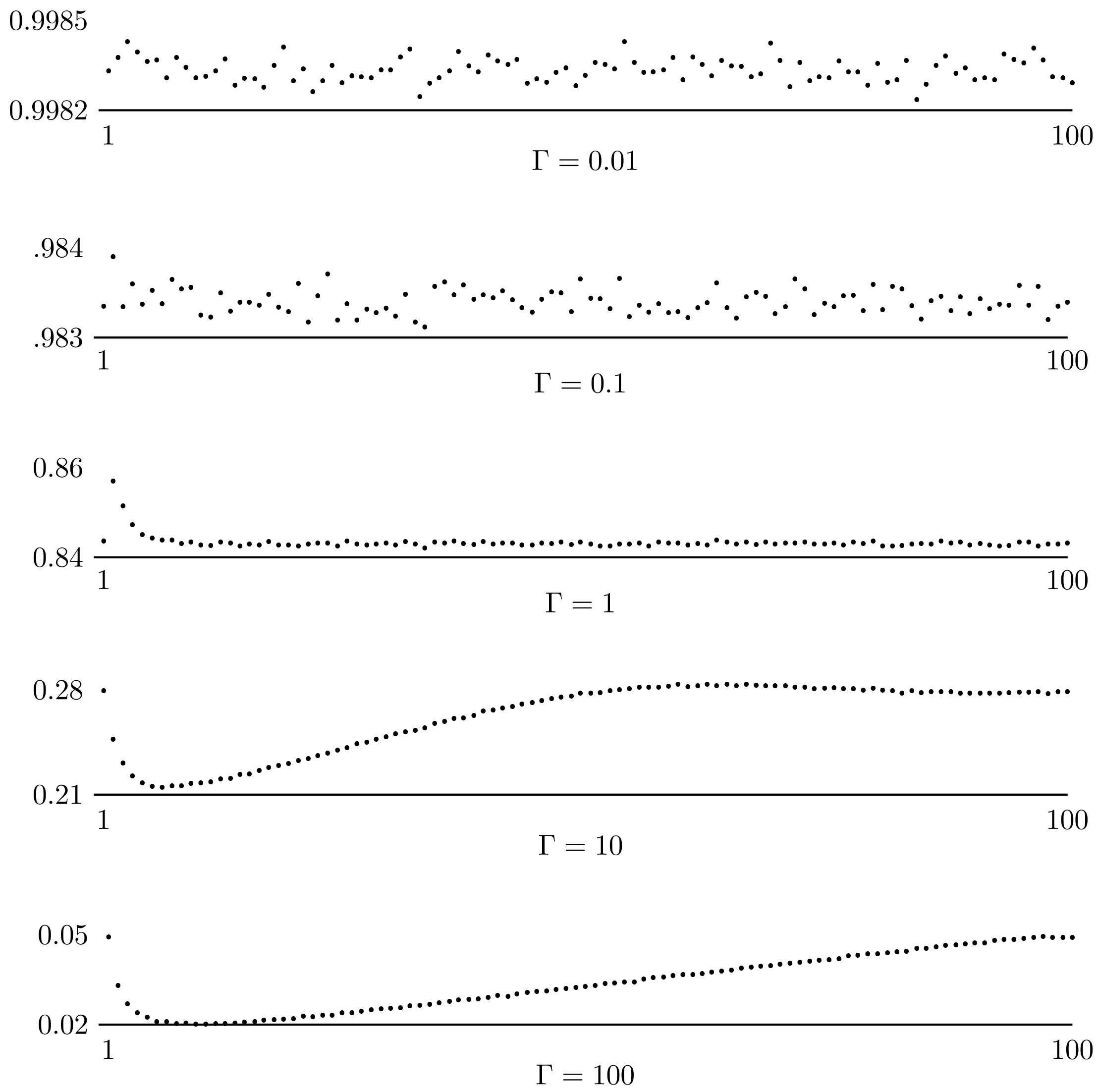

Figure 14: (E, U): the empirical hitting ratio

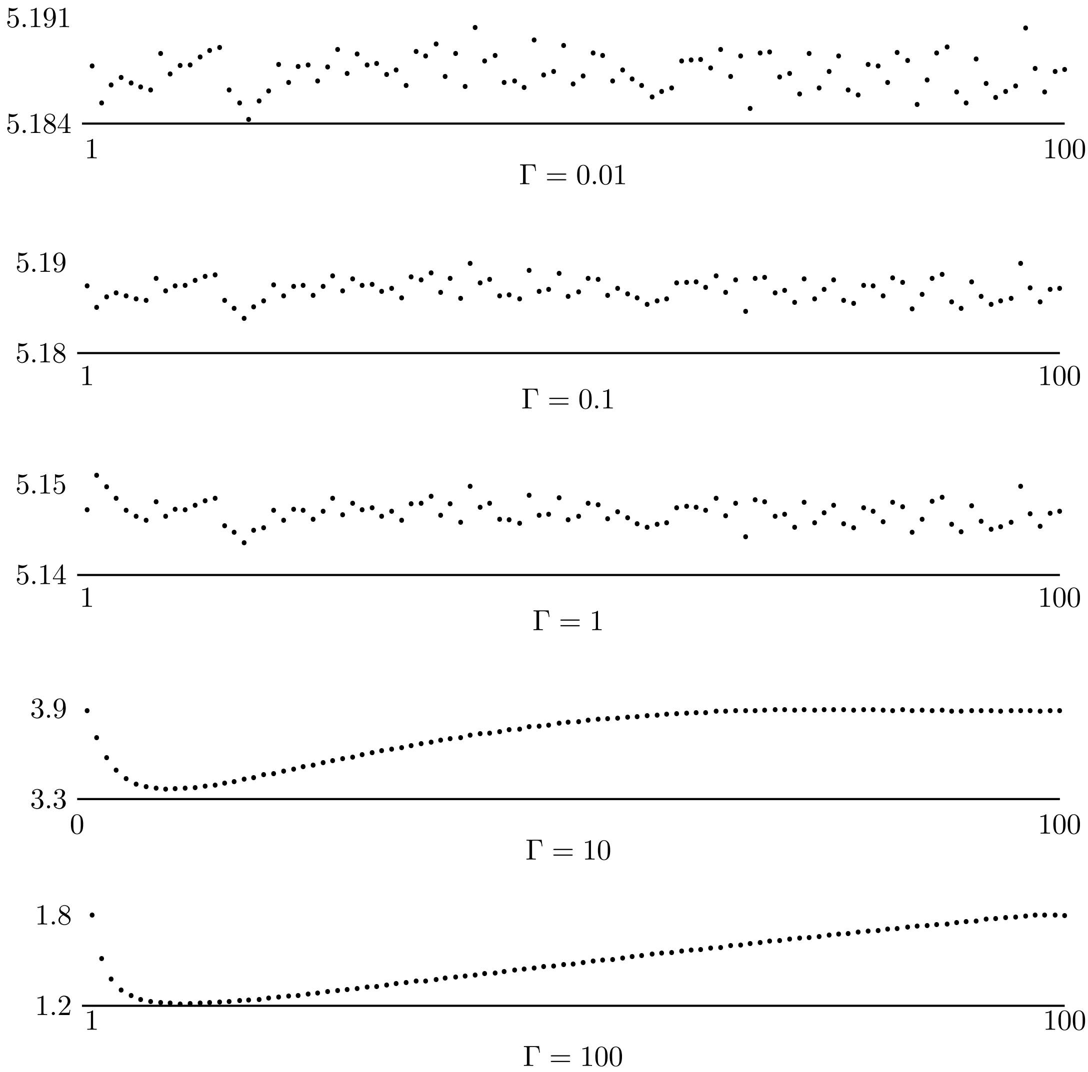

Figure 15: (E, U): the empirical expected result

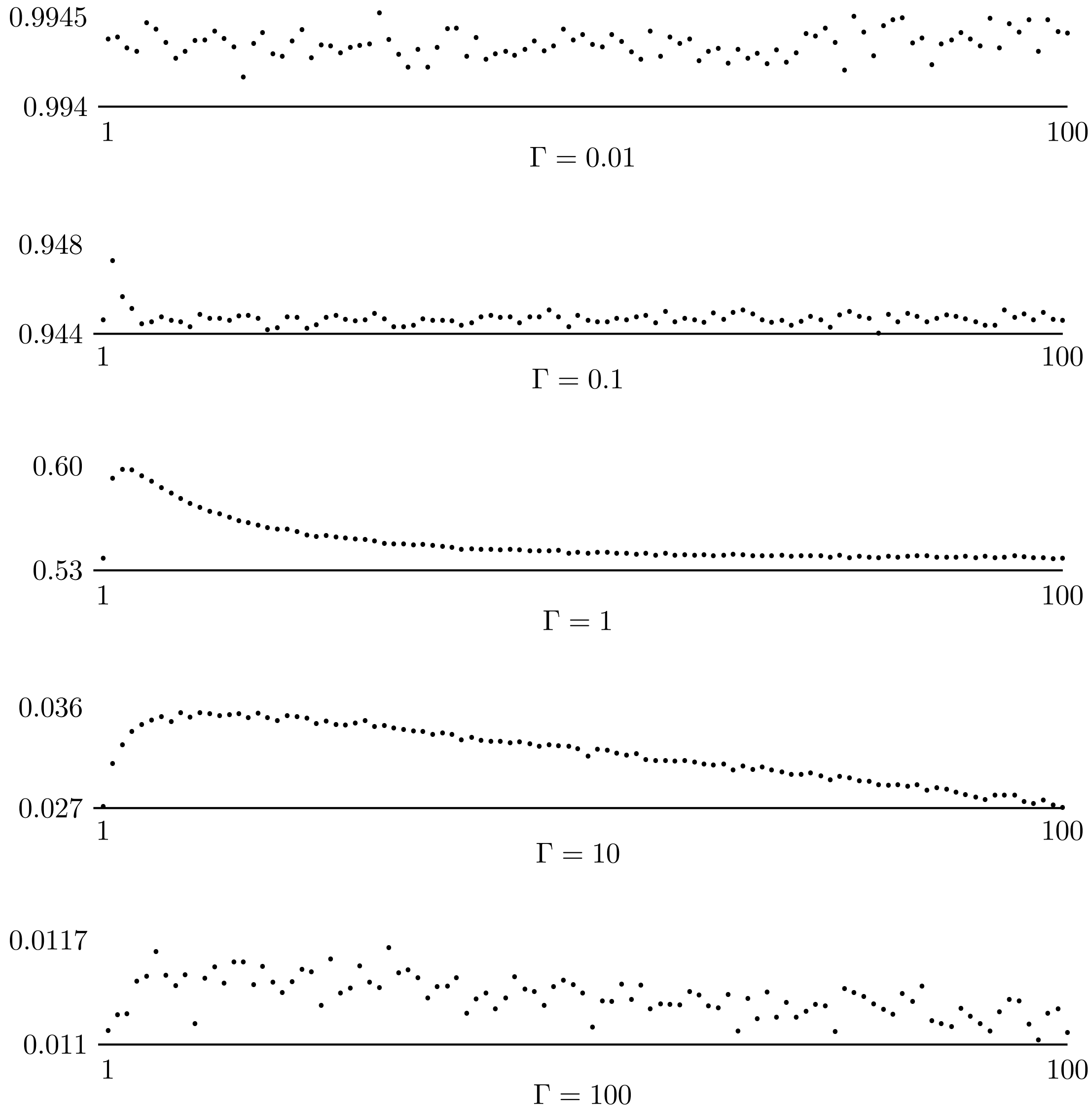

Figure 16: (E, N): the empirical hitting ratio

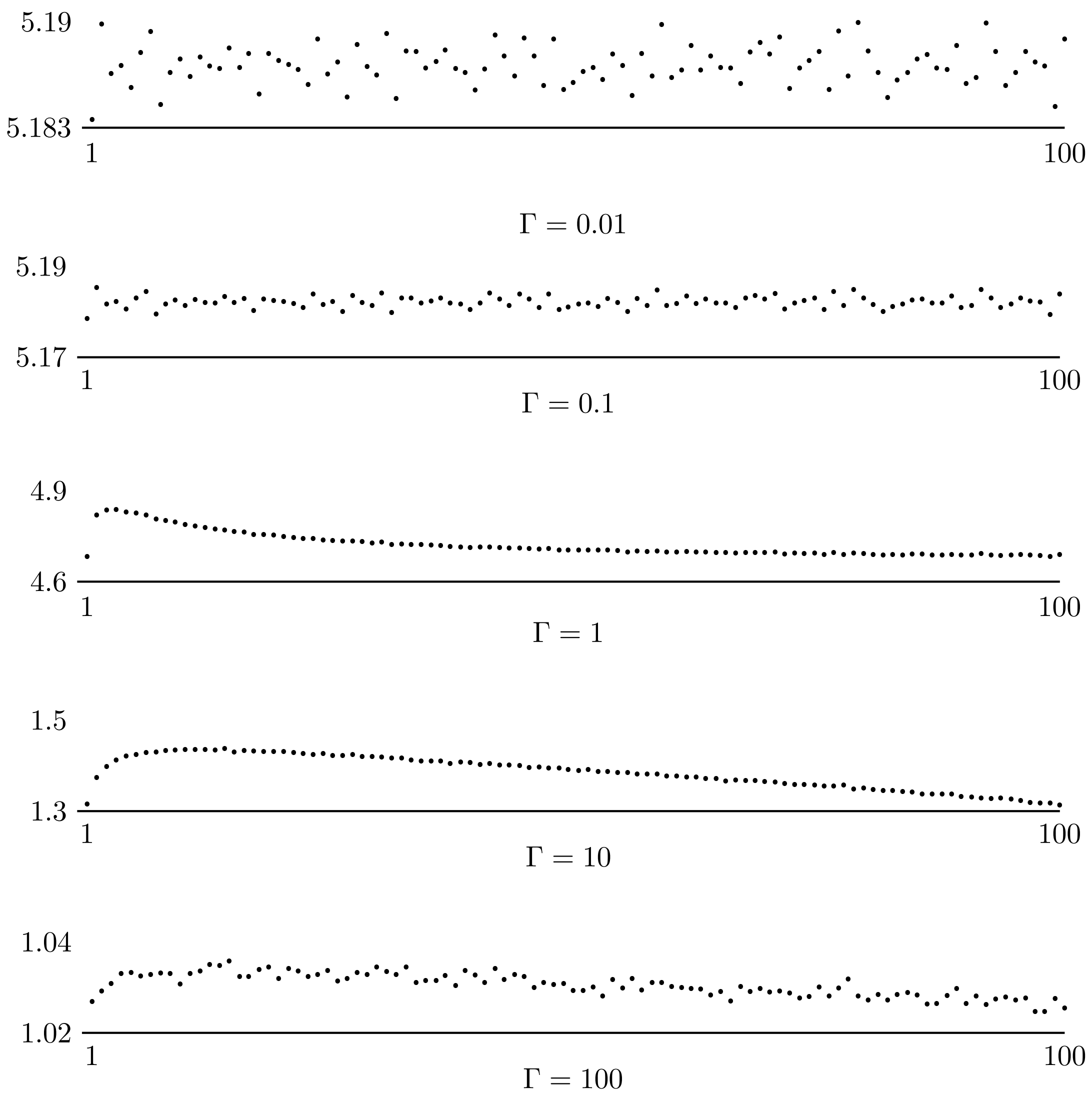

Figure 17: (E, N): the empirical expected result

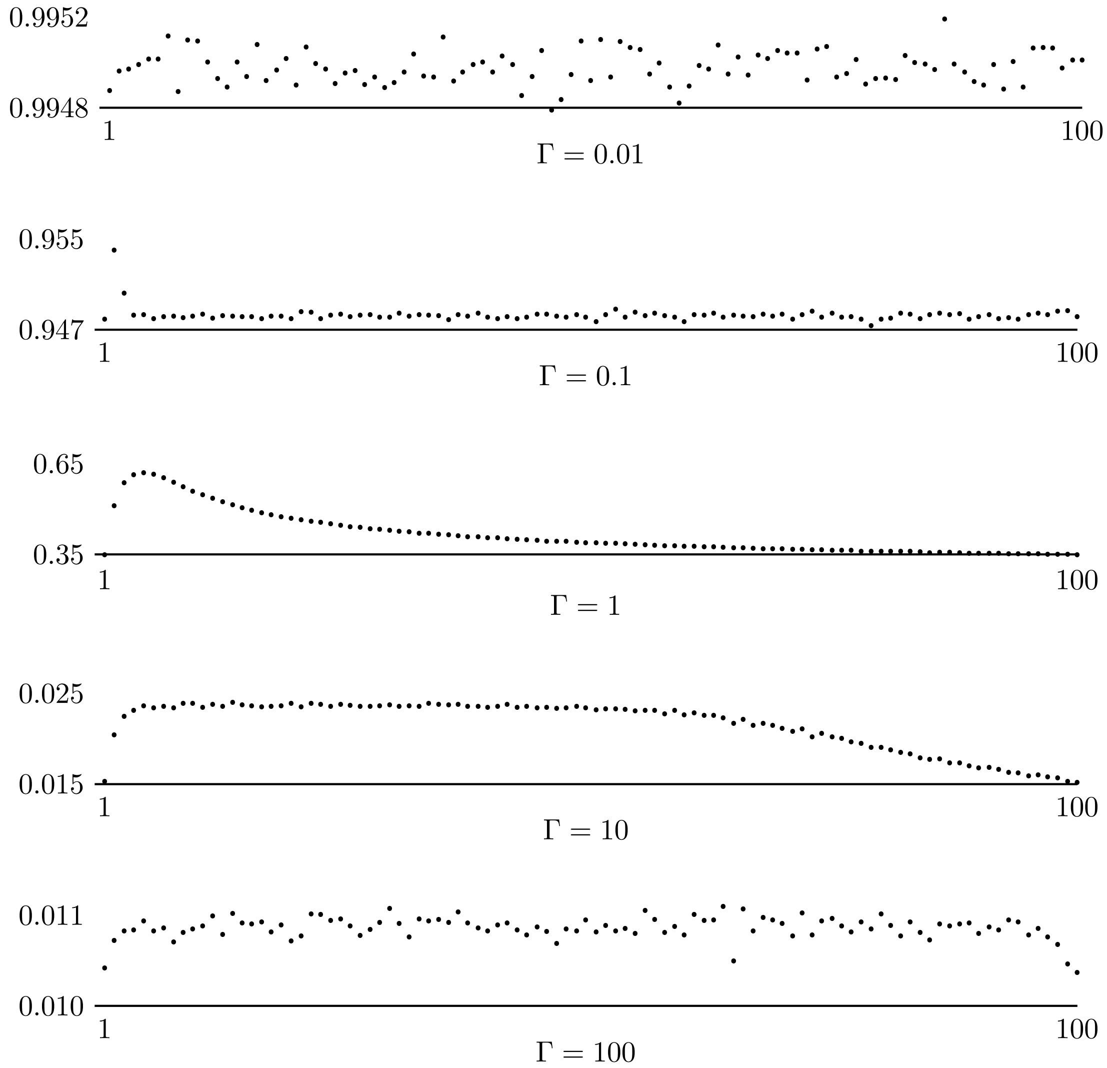

Figure 18: (E, E): the empirical hitting ratio

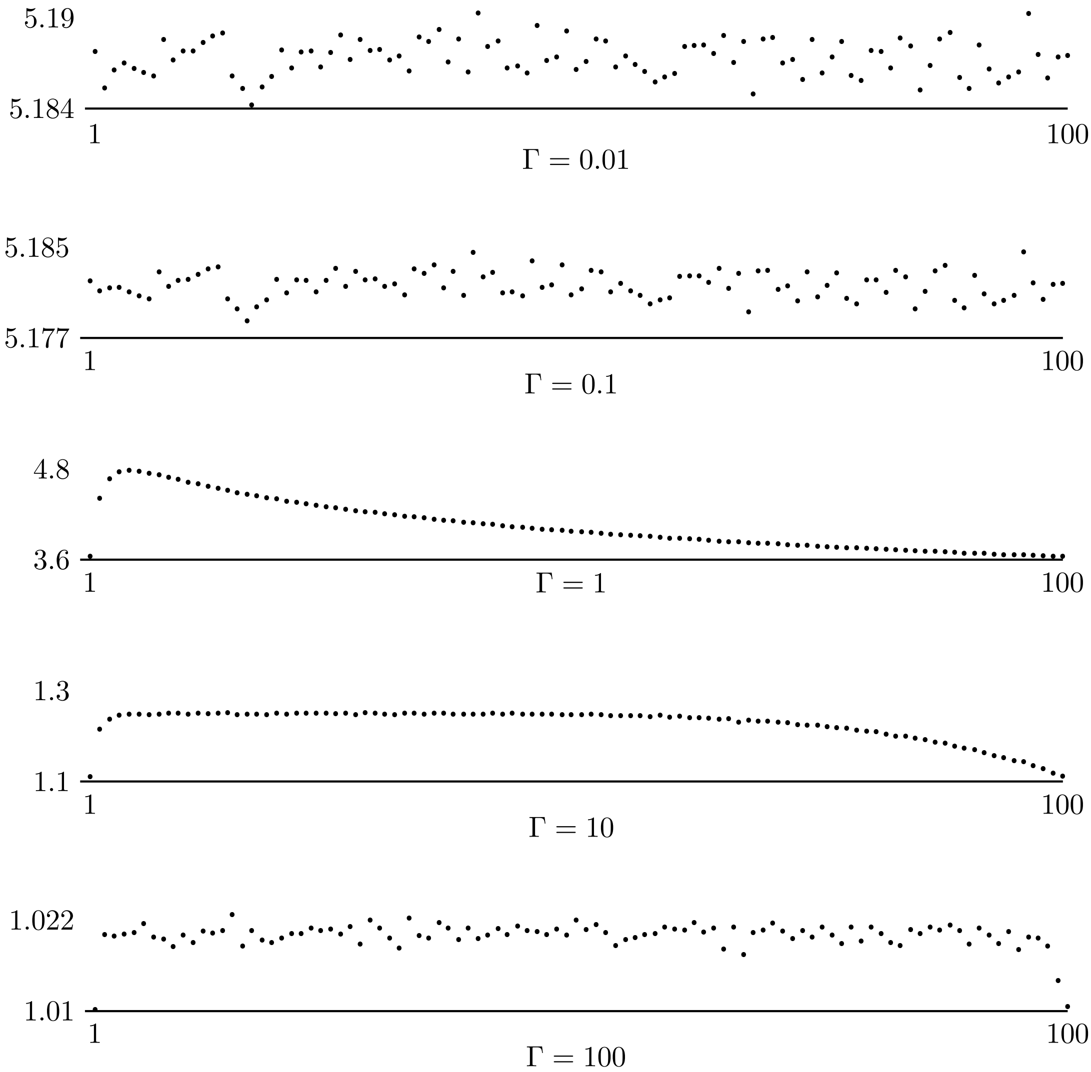

Figure 19: (E, E): the empirical expected result