Société de Calcul Mathématique SA
Mathematical Modelling Company, Corp
Tools for decision help
since 1995

Simple Random Walks in the Plane:

An Energy-Based Approach

Part I : Basic Facts

Bernard Beauzamy
August 2017

I. Introduction

We consider a simple random walk in the plane : a sequence of random variables X_n with values $\pm 1,$ probability $1/2$ in each case. Let $S_N = \sum_{n=1}^{N} X_n$ be the sum of the first N variables. This random walk can be viewed as a game between two players A and B; at the n^{th} step, the first player receives 1 Euro from the second player if $X_n = +1$ and conversely if $X_n = 1.$ So the sum S_N represents the increase of fortune of A compared to B at the end of N games ; this increase may of course be positive or negative. At the initial moment, we set $S_0 = 0.$ Besides that, each player has an initial fortune, which may be finite or infinite. Depending on the settings, the game may stop when one of the players is ruined (his fortune is equal to 0). The general question is to study the behavior of S_N (possible values, with their probabilities) and their asymptotic distribution, when $N \to +\infty,$ assuming possible limitations on each fortune, or on both.

Among the many existing results on this topic, let us mention in particular:

- Feller's "Gambler's ruin" ; see [Feller]. The problem may be stated as follows : given an initial fortune and a barrier, what is the probability to reach the barrier without having first reached the barrier $y=0$ (which means ruin) ? The gambler's ruin does not care about a specific time, whereas we compute the energy on each vertical for each specific time. We thank Doron Zeilberger for useful discussions about this comparison.

- Khintchin's law of the iterated logarith (1924); see [Khintchin]: almost surely, when $n \to +\infty$:

$$\limsup \frac{S_n}{\sqrt{2nLog\left(Log\left(n\right)\right)}} = +1 \text{ and } \liminf \frac{S_n}{\sqrt{2nLog\left(Log\left(n\right)\right)}} = -1$$

We will here present a new approach to such problems, which is "energy based" and not probabilistic in nature. This will allow us to develop a unified framework, and to obtain quantitative estimates which were not known previously.

Indeed, the probabilistic appearance of Khinchin's laws is misleading. Looking at such a statement, everyone has the impression that, for a given player, there are some unknown forces which will, sooner or later, bring his fortune close to Khintchin's curves (a Khintchin curve is of the form $y = \pm\sqrt{2x\, Log\left(Log\left(x\right)\right)}$, of course). This is completely wrong ; at any time, the game is only governed by the ± 1 rule, with equal probability.

What Khintchin's laws say, and, more generally, what any result about random walks says, is that there are more paths with some properties than paths with other properties. They are not individual results about each path; they are results about the number of paths with a given characteristic. Such results are in fact of combinatorial nature. For instance, at time n, the proportion of paths which never touched the curve $y=\sqrt{n}$ tends to 0 when $n \to +\infty$.

Our approach relies upon a concept derived from "energy absorption". Our aim is to obtain quantitative estimates, of the following form:

Given a curve $y=\varphi(n)$, or possibly a couple of symmetric curves $y=\pm\varphi(n)$, what is the proportion of paths which never touched the curve(s) before the instant n ?

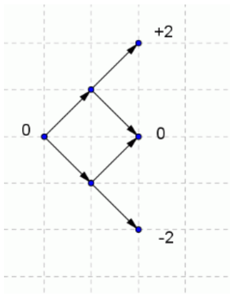
II. Basic settings

We consider that, at time 0, a unit of energy is put at the origin. This unit will then divide itself in two halves, at time $n=1$, one at the point $(1,1)$ and one at the point $(1,-1)$. More generally, every time a division point is met, the available energy divides equally into the two possible paths. So, for instance, at the time $n=2$, 3 points will receive some energy, namely $(2,2)$ receives 1/4, $(2,0)$ receives 1/2, $(2,2)$ receives 1/4. At any step, in this configuration, the sum is always 1.

We observe that, obviously, at any time n, we have $|S_n| \leq n$.

The values of S_n are even if n is even, and are odd if n is odd. In what follows, we will almost always restrict ourselves to the case where n is even. This means that the elementary game consists in two repetitions, $X_1 + X_2$, with :

$$P(X_1 + X_2 = -2) = \frac{1}{4}, P(X_1 + X_2 = 0) = \frac{1}{2}, P(X_1 + X_2 = 2) = \frac{1}{4} \qquad (1)$$



So an energy put at any point will divide into four : one fourth 2 steps above, one half at the same level, one fourth 2 steps below: see picture.

The following Lemma is well-known (see for instance [1]) ; it simply reflects the combinatorics:

Lemma 1. - *Let $A_{2n,2k}$ be the point of coordinates $(2n,2k)$, with $k=-n,\dots,n$. The number of paths from 0 to $A_{2n,2k}$ is:*

$$N(2n,2k) = \binom{2n}{n+k}$$

Proof of Lemma 1

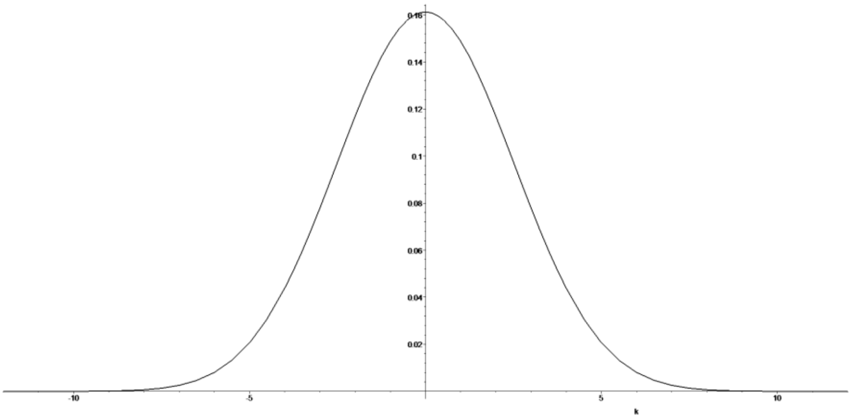
If we want to reach this point in $2n$ steps, we need x times the value 1 and y times the value -1, with $x+y=2n$ and $x-y=2k$, which gives $x=n+k,\ \ y=n-k$. So there are $\binom{2n}{x}$ possible paths, which proves the result.

We write $N(A) = N(0 \to A)$ for the total number of paths, starting at 0, finishing at A and, more generally, $N(A \to B)$ for the number of paths starting at A, finishing at B.

In this basic setting, since the energy 1 is put at O and since there is a total of 2^{2n} possible paths $N(A_{2n,2k})$ at time $2n$, each point $A_{2n,2k}$ receives an amount of energy equal to:

$$P(S_{2n} = 2k) = e(A_{2n,2k}) = \frac{1}{2^{2n}}\binom{2n}{n+k} \qquad (2)$$

We see that the repartition of energy is given by a binomial law: there is more energy at the central points and very little energy at the extreme points ($A_{2n,2n}$ and $A_{2n,-2n}$).



Example of energy distribution for $n=12$

We consider, for simplicity, only the even values of n,k ; let $f(n,k)=e\left(A_{2n,2k}\right)$ be the energy put at the point of coordinates $2n,2k$. It satisfies for any k :

$$f(n,k)=\frac{1}{4}f(n-1,k-1)+\frac{1}{2}f(n,k)+\frac{1}{4}f(n,k+1) \tag{1}$$

Now, we can restrict ourselves to $k\geq 0$, using the symmetry of the process. Since $f(n,1)=f(n,-1)$, we can write :

$$f(n,0)=\frac{1}{2}f(n-1,0)+\frac{1}{2}f(n-1,1) \tag{2}$$

and (1) remains valid for $k>0$.

Proposition 2. - *For all* $k\geq 0$, $f(n,k)\to 0$ *when* $n\to+\infty$.

Proof of Proposition 2

We consider the Banach space l_1 of sequences x_k such that $\sum_{k=0}^{+\infty}|x_k|<+\infty$. The linear operator T which sends a sequence x_k to the sequence y_k defined by:

$$y_0=\frac{1}{2}x_0+\frac{1}{2}x_1$$

and, for $k\geq 1$:

$$y_k=\frac{1}{4}x_{k-1}+\frac{1}{2}x_k+\frac{1}{4}x_{k+1}$$

is an isometry on the positive sequences (the sequence x_k and the sequence y_k have same sum). One checks immediately that if the x_k are decreasing, so are the y_k. We start with the sequence

$X_0 = (1, 0, \ldots)$ and consider its iterates under the operator T; let $X_n = T^n X_0$. Let $X_n(0)$ be the first coordinate of the sequence X_n. It is clear on the definition that the sequence $X_n(0)$ is decreasing, and the same applies to each coordinate $X_n(j)$. Since all are positive, each sequence $X_n(j)$ must have a limit, denoted by l_j, when $n \to +\infty$. But we have:

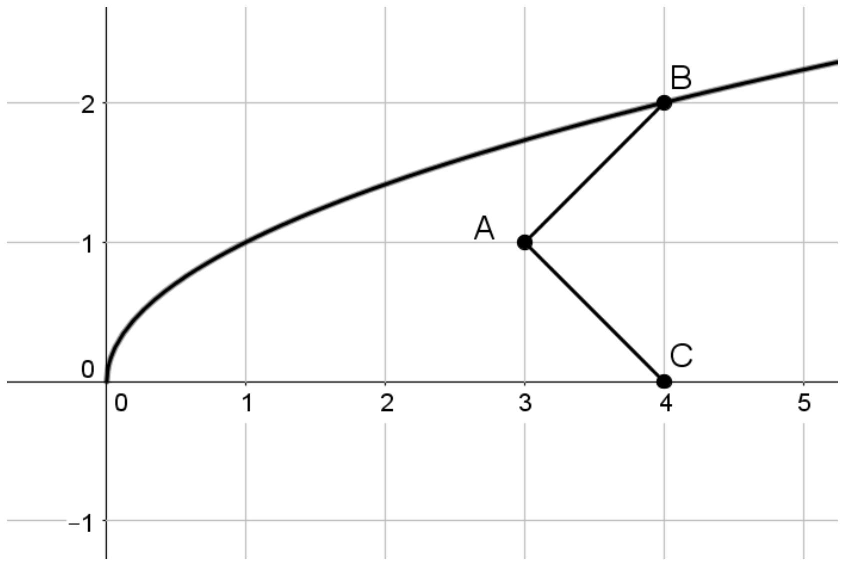
$$X_n(0) = \frac{1}{2} X_{n-1}(0) + \frac{1}{2} X_{n-1}(1)$$

which implies $l_0 = \frac{1}{2} l_0 + \frac{1}{2} l_1$ and therefore $l_0 = l_1$. The same applies to each coordinate: $l_0 = l_1 = \ldots = l_j$. The limiting value must be the same for all coordinates. But we have the condition $\sum_j X_n(j) = 1$, for any n, which implies $\sum_j l_j \le 1$, so the l_j must all be 0. This proves the Proposition.

In this preliminary approach, the total amount of energy remains the same at each time step.

Now, we introduce a curve, $y = \varphi(x)$ located in the upper half-plane (the same holds for the lower half-plane, of course), and we want to investigate the probability that the random walk, up to time n, remains constantly below this curve (which means that $S(j) < \varphi(j)$ for all $j = 1, \ldots, n$. In some cases, we will investigate the probability to remain between the curve and its symmetric, which means $|S(j)| < \varphi(j)$.

Our representation, in order to investigate this phenomenon, will be the fact that the curve φ absorbs the energy. This means that, for any path which touches the curve, the corresponding energy disappears.



Example of energy absorption

In this example, the point A sends its energy to both B and C, but B is on the curve we have introduced, so this part of the energy disappears, and we are left with $e(C)=\frac{1}{2}e(A)$.

The curve we introduce will be called the critical curve. It may be considered as a "black frontier" (in the sense of a black hole), meaning that it absorbs all energy it receives, and sends back nothing.

We have:

Proposition 3. - *Let $y=\varphi(x)$ be any critical curve, in the upper half-plane. The total energy left, at time n, is equal to the total probability to reach any of the points $A_{n,k}$ below the curve, that is $k<\varphi(n)$, without ever touching the curve at any time before ($j \le n$).*

Proof of Proposition 3

This is a mere rephrasing of the disappearance of energy. Any time a path touches the curve, it is annihilated, so what remains is the set of paths which never touched the curve.

If a time n is fixed, and a curve φ is fixed, we will call admissible a path with never touches it (at any time $j \le n$). For any point A in the plane, let $N_{ad}(A)$ be the number of admissible paths which reach A, and $p_{ad}(A)=\frac{N_{ad}(A)}{2^n}$ the probability to reach A by an admissible path. Proposition 2 states that:

$$\sum_{k=-n}^{n} e(A_{n,k}) = \sum_{k<\varphi(n)} p_{ad}(A_{n,k})$$

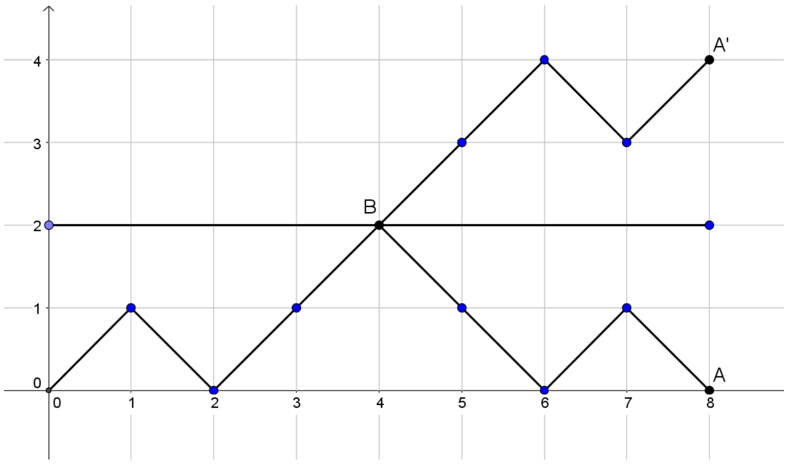
III. The case of an horizontal line

We now compute the number of admissible paths when the critical curve is a simple horizontal line segment. There are some differences, depending if y is odd or even; we will restrict ourselves to the odd case.

Lemma 4. - *Let $y=2\xi+1$ ($\xi \ge 0$) be an horizontal line segment. Let $A_{2n,2k}$, with coordinates $(2n,2k)$, be any point that the random walk may reach, with $k \le \xi$. The number of paths, starting at 0, finishing at $A_{2n,2k}$, which touch the horizontal segment at a time before $2n$ is $N(A_{2n,2\xi+2-2k})$, where $A_{2n,2\xi+2-2k}$ is the symmetric of $A_{2n,2k}$ with respect to the line segment.*

Proof of Lemma 4

This property is well-known (see for instance [1]), under the name of "reflexion principle" :



The reflexion principle

Let B be the first time a path touches the segment (there may be several). There are as many paths from B to A than from B to A', symmetric of A with respect to the barrier.

The symmetric of $A_{2n,2k}$ is $A_{2n,4\xi+2-2k}$. So the number of paths which touch the segment $y=2\xi+1$ at any time before n is, by Lemma 1 :

$$N\left(A_{2n,4\xi-2k}\right)=\binom{2n}{n+2\xi+1-k}$$

Therefore, the number of paths which reach $A_{2n,2k}$ without ever touching the segment $y=2\xi+1$ is:

$$N_{ad}\left(A_{2n,2k}\right)=\binom{2n}{n+k}-\binom{2n}{n+2\xi+1-k}$$

This proves Lemma 4.

In the sequel, we denote by W_{2n} the "vertical" at time $2n$. This is the set of points $A_{2n,2k}$, $k=-n,\dots,n$. We also denote by E_{2n} the total energy on this vertical : $E_{2n}=\sum_{k=-n}^{n} e\left(A_{2n,2k}\right)$.

Proposition 5. - *Assume that our critical curve is the line segment $y=2\xi+1$. The energy left at time $2n$ is:*

$$E_{2n}=1-\frac{2}{2^{2n}}\sum_{j=\xi+1}^{n}\binom{2n}{n+j}=\frac{1}{2^{2n}}\sum_{j=-\xi}^{\xi}\binom{2n}{n+j}$$

Proof of Proposition 5

The critical line segment $y=2\xi+1$ has two effects :

– No point $A_{2n,2k}$ above this segment receives any energy at all ; there is a drop of total energy equal to the probability to reach this point;

– For every point strictly below this segment, there is a drop of energy equal to the probability to reach its symmetric.

Since both terms are equal, the total drop of energy (that is the total energy "swallowed" by the segment), instead of reaching W_{2n}, is $\frac{2}{2^{2n}} \sum_{j=\xi+1}^{n} \binom{2n}{n+j}$. This proves Proposition 5.

It is clear from Proposition 5 that $E_{2n} \to 0$ when $n \to +\infty$. Indeed, in the expression $E_{2n} = \frac{1}{2^{2n}} \sum_{j=-\xi}^{\xi} \binom{2n}{n+j}$, there is a fixed number of terms and each term tends to 0 when $n \to +\infty$. But we want to make this statement quantitative.

We have, using the approximation of the binomial law by the normal law:

$$\frac{1}{2^{2n}} \sum_{j=-\xi}^{\xi} \binom{2n}{n+j} \approx \int_{-\xi}^{\xi} \exp\left(-\frac{t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}}, \quad \text{with } \sigma^2 = 2n.$$

In order to make this approximation precise, we use Berry-Esseen Theorem [Berry-Esseen], which may be stated as follows:

For all x and all n:

$$\left| P\left(\frac{S_n}{\sqrt{n}} \le x\right) - \int_{-\infty}^{x} \exp\left(\frac{-t^2}{2}\right) \frac{dt}{\sqrt{2\pi}} \right| \le \frac{1}{\sqrt{n}}$$

which we write under the form:

$$\left| P\left(S_{2n} \le x\sqrt{2n}\right) - \int_{-\infty}^{x\sqrt{2n}} \exp\left(\frac{-t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}} \right| \le \frac{1}{\sqrt{2n}}, \text{ with } \sigma = \sqrt{2n}$$

or, with $\xi = \frac{x\sqrt{2n}}{2}$:

$$\left| P\left(S_{2n} \le 2\xi\right) - \int_{-\infty}^{2\xi} \exp\left(\frac{-t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}} \right| \le \frac{1}{\sqrt{2n}}$$

But, by (2):

$$P\left(S_{2n} \le 2\xi\right) = \frac{1}{2^{2n}} \sum_{k \le \xi} \binom{2n}{n+k}$$

Therefore:

$$\left| \frac{1}{2^{2n}} \sum_{k \le \xi} \binom{2n}{n+k} - \int_{-\infty}^{2\xi} \exp\left(\frac{-t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}} \right| \le \frac{1}{\sqrt{2n}} \qquad (3)$$

and also with $-\xi$:

$$\left| \frac{1}{2^{2n}} \sum_{k \le -\xi} \binom{2n}{n+k} - \int_{-\infty}^{-2\xi} \exp\left(\frac{-t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}} \right| \le \frac{1}{\sqrt{2n}}$$

Taking the difference, we obtain:

$$\left| \frac{1}{2^{2n}} \sum_{k=-\xi}^{\xi} \binom{2n}{n+k} - \int_{-2\xi}^{2\xi} \exp\left(\frac{-t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}} \right| \le \sqrt{\frac{2}{n}} \qquad (4)$$

From (4), we immediately deduce an estimate for the sum $\frac{1}{2^{2n}} \sum_{k=-\xi}^{\xi} \binom{2n}{n+k}$, valid for all n :

Proposition 6. - *For all ξ and n, we have the estimate:*

$$E_{2n} \le \frac{2\xi}{\sqrt{\pi n}} + \sqrt{\frac{2}{n}}$$

Proof of Proposition 6

Indeed, by (4):

$$E_{2n} \le \int_{-2\xi}^{2\xi} \exp\left(\frac{-t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}} + \sqrt{\frac{2}{n}} = \int_{-\frac{2\xi}{\sqrt{2n}}}^{\frac{2\xi}{\sqrt{2n}}} \exp\left(\frac{-t^2}{2}\right) \frac{dt}{\sqrt{2\pi}} + \sqrt{\frac{2}{n}} \le \frac{2\xi}{\sqrt{\pi n}} + \sqrt{\frac{2}{n}}$$

which proves Proposition 6.

For $\xi = 0$, we find the estimate $E_{2n} \le \sqrt{\frac{2}{n}}$, whereas a direct application of Stirling's formula gives $E_{2n} \le \frac{1}{\sqrt{\pi n}}$, so the estimate in Proposition 6 is not best possible.

IV. Gaussian interpretation

Formula (3) gives immediately, with $\sigma^2 = 2n$

$$\left| \frac{1}{2^{2n}} \sum_{k=\xi_1}^{\xi_2} \binom{2n}{n+k} - \int_{2\xi_1}^{2\xi_2} \exp\left(\frac{-t^2}{2\sigma^2}\right) \frac{dt}{\sigma\sqrt{2\pi}} \right| \le \sqrt{\frac{2}{n}} \qquad (1)$$

This formula has an interpretation, namely that the energy, on any vertical W_{2n}, between the levels $2\xi_1$ and $2\xi_2$, may be viewed as a gaussian integral between these two levels, with a Gaussian law, the variance of which is the distance between 0 and the vertical (this distance is $2n$). The error in this approximation is smaller than $\sqrt{\frac{2}{n}}$.

This setting is much easier to handle, since Gaussian integrals are simpler to manipulate than binomial sums. Let us give a complete reinterpretation of the previous paragraph: energy absorption in case of a barrier at ξ.

In this continuous setting, there is no need to differentiate between the odd and even cases, which is also a simplification.

The symmetric of a point $A_{n,t}$ with respect to the barrier $y = \xi$ is $A_{n,2\xi-t}$. Therefore, the density of energy sent by 0 to the point $A_{n,t}$, taking into account the annihilation by the barrier, is :

$$f_n(t) = \frac{1}{\sigma\sqrt{2\pi}} \left(\exp\left(-\frac{t^2}{2\sigma^2}\right) - \exp\left(-\frac{(2\xi - t)^2}{2\sigma^2}\right) \right), \text{ for } t \le \xi \text{ , 0 if } t > \xi \qquad (2)$$

with $\sigma = \sqrt{n}$. So this is simply the difference of two gaussian functions with same variance.

From (2), we easily obtain the profile of energy, on the vertical W_n, that is the position of the point of maximal energy:

Proposition 7. - *The point of maximal energy on the vertical* W_n *has coordinate:*

$$t_n \approx \frac{3\xi}{2} - \frac{1}{2}\sqrt{\xi^2 + 4n}$$

Proof of Proposition 7

Indeed, we have to find the maximum of the function f_n. But:

$$f_n'(t) = \frac{1}{\sigma\sqrt{2\pi}}\left(-\frac{t}{\sigma^2}\exp\left(-\frac{t^2}{2\sigma^2}\right)+\frac{t-2\xi}{\sigma^2}\exp\left(-\frac{(t-2\xi)^2}{2\sigma^2}\right)\right)$$

So, the condition $f_n' = 0$ is equivalent to:

$$\frac{t}{t-2\xi} = \exp\left(\frac{t^2}{2\sigma^2}-\frac{(t-2\xi)^2}{2\sigma^2}\right) = \exp\left(\frac{-2\xi(\xi-t)}{n}\right) \tag{3}$$

Consider first the function $h(t) = \frac{t}{t-2\xi}$; the derivative is $h'(t) = \frac{-2\xi}{(t-2\xi)^2} < 0$, so the function is decreasing, has the limit 1 at $-\infty$ and takes the value -1 for $t = \xi$. The function $g(t) = \exp\left(\frac{-2\xi(\xi-t)}{n}\right)$ is increasing, has limit 0 when $t \to -\infty$ and takes the value 1 for $t = \xi$. Therefore, a unique solution t_n of equation (3) exists. When $n \to +\infty$, we have the rough estimate:

$$\frac{t}{t-2\xi} \sim 1 - \frac{2\xi(\xi-t)}{n}$$

that is:

$$t_n \approx \frac{3\xi}{2} - \frac{1}{2}\sqrt{\xi^2+4n}$$

which implies that $t_n \to -\infty$ and proves Proposition 7.

V. References

[Berry-Esseen]
https://en.wikipedia.org/wiki/Berry%E2%80%93Esseen_theorem

[Feller] W. Feller : An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Edition, Wiley series in Probabilities.

[Khinchine] A. Khinchine. "Über einen Satz der Wahrscheinlichkeitsrechnung", Fundamenta Mathematica, 6:9-20, 1924.