Handout Limit Theorems Probability Theory 2MBS10

Remco van der Hofstad*

June 7, 2024

Abstract

In this handout we discuss some important limit theorems with their proofs. Our main focus is on the following aspects: (a) We define what *convergence in distribution* means for a sequence of random variables; (b) We give a bound on some probabilities, in particular the *Markov* and *Chebychev* inequalities; (c) We discuss the *law of large numbers*, and provide a proof; (d) Finally, we discuss the central limit theorem (CLT), with proof and many applications of this theorem. Because of its importance, the central limit theorem is sometimes called the *Fundamental Theorem of Statistics*. We give its (self-contained) proof since we find it important that any student with a bachelor in mathematics has seen the proof of this theorem. The proof itself is not part of the course material, the applications of the CLT are. We provide a simple algorithm to apply the CLT that guarantees a successful application of it, and we encourage students to follow this four-step algorithm.

1 Convergence in Distribution

In this section, we define *convergence in distribution* for sequences of random variables, and provide some examples. Recall that a random variable is defined as a function $X : \Omega \to \Omega$, where Ω denotes the state space. A sequence of random variables $(X_n)_{n\geq 1}$ is therefore a sequence of functions, and there are different ways in which a sequence of functions can converge. In this handout, we only discuss *convergence in distribution*:

Definition 1.1 (Convergence in distribution)**.** *A sequence of random variables $(X_n)_{n\geq 1}$ with distribution functions $(F_{X_n})_{n\geq 1}$ converges in distribution to a random variable X with distribution function F_X if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{1}$$

for all $x \in \mathbb{R}$ where $x \mapsto F_X(x)$ is continuous in x. We write this as $X_n \xrightarrow{d} X$.

Convergence in distribution means that probabilities with respect to the random variables X_n converge to the corresponding probability for X. In particular, this definition implies that

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x) \tag{2}$$

*Department of Mathematics and Computer Science, Eindhoven University of Technology, Box 513, 5600 MB Eindhoven, The Netherlands; rhofstad@win.tue.nl

for all values of x where the distribution function is continuous. These are called *continuity points* of the distribution. Since

$$\lim_{h \searrow 0} \mathbb{P}(X \leq x) - \mathbb{P}(X \leq x - h) = \mathbb{P}(X = x), \tag{3}$$

the continuity points are precisely those for which $\mathbb{P}(X = x) = 0$.

The restriction to values of x where F_X is continuous is necessary, as it is not clear what happens in the discontinuity points of F_X. We next provide an example of this:

Example 1.2 (A random variable that converges in distribution, but not in the discontinuous point)**.** *Let $Y_n \sim \mathsf{Bin}(n, 1/2)$, and $X_n = Y_n/n$. Based on Section 3 below, we see that for $x < 1/2$,*

$$\mathbb{P}(X_n \leq x) \to 0, \tag{4}$$

while for $x > 1/2$,

$$\mathbb{P}(X_n \leq x) \to 1. \tag{5}$$

This implies that $X_n \xrightarrow{d} 1/2$. We know that $X = 1/2$ is a (deterministic) random variable, with $F_X(x) = 0$ for $x < 1/2$ and $F_X(x) = 1$ for $x \geq 1/2$. In particular, we have $F_X(1/2) = 1$. However, if n is odd and $x = 1/2$, then $\mathbb{P}(X_n/n \leq 1/2) = 1/2$ due to symmetry. So, $F_{X_n}(1/2)$ does converge, but not to $F_X(1/2)$. The definition of convergence in distribution in Definition 1.1 takes this into account.

In the special case where X is a *continuous* random variable, $x \mapsto F_X(x)$ is continuous in every value of x, which implies that $F_{X_n}(x) \to F_X(x)$ for *all* $x \in \mathbb{R}$. We provide some examples:

Example 1.3 (Convergence of a binomial to a Poisson)**.** *During the lectures, we already saw that if $X_n \sim \mathsf{Bin}(n, \lambda/n)$ and $X \sim \mathsf{Poi}(\lambda)$, the following holds: $\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$. This is not exactly the same as convergence in distribution, but it does imply it. This can be seen by noting that*

$$F_{X_n}(x) = \sum_{k \leq x} \mathbb{P}(X_n = k), \tag{6}$$

and, by assumption, the (finitely-many) summand converge. The reader is asked to verify this in detail in the following exercise:

Exercise 1.1 (Convergence of integer-valued random variables)**.** *Show that convergence in distribution is equivalent to $\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$ when both $(X_n)_{n\geq 1}$ and X are integer-valued random variables, and conclude that $X_n \xrightarrow{d} X$ when $X_n \sim \mathsf{Bin}(n, \lambda/n)$ and $X \sim \mathsf{Poi}(\lambda)$.*

Example 1.4 (Minima for uniform random variables)**.** *Let $U_1, \ldots, U_n \sim U[0, a]$, with $a > 0$ be independent and identically distributed (i.i.d.). Then we can show that $X_n = n \min\{U_1, \ldots, U_n\} \xrightarrow{d} E$, with $E \sim \mathsf{Exp}(1/a)$. Indeed, for every $x > 0$,*

$$\begin{aligned} \mathbb{P}(X_n \leq x) &= 1 - \mathbb{P}(X_n > x) = 1 - \mathbb{P}(U_1 > x/n, \ldots, U_n > x/n) = 1 - (1 - x/(na))^n \\ &\to 1 - \mathrm{e}^{-x/a}. \end{aligned} \tag{7}$$

Hence, we conclude $X_n \xrightarrow{d} E$ with $E \sim \mathsf{Exp}(1/a)$.

Exercise 1.2 (Convergence of a geometric to an exponential)**.** *Define $X_n \sim \mathsf{Geo}(1/n)$. Show that $X_n/n \xrightarrow{d} X$, where $X \sim \mathsf{Exp}(1)$.*

2 Bounds on Probabilities

In this section, we provide bounds on some probabilities. These bounds can be used to show that random variables converge to a degenerate (deterministic) random variable. In the next section we use these results to prove the weak law of the large numbers. We start with *Markov's inequality:*

Theorem 2.1 (Markov's inequality)**.** *Let X be a non-negative random variable with $\mathbb{E}[X] < \infty$, and $a > 0$. Then*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{8}$$

Proof. We can bound, since $X \geq 0$,

$$\mathbb{E}[X] \geq \mathbb{E}[X \mathbb{1}_{\{X \geq a\}}] \geq a\mathbb{E}[\mathbb{1}_{\{X \geq a\}}] = a\mathbb{P}(X \geq a). \tag{9}$$

Here, $\mathbb{1}_{\{X \geq a\}}$ denotes the indicator of the event $\{X \geq a\}$. Division by $a > 0$ provides the claim. □

A similar inequality is *Chebychev's inequality:*

Theorem 2.2 (Chebychev's inequality)**.** *Let X be a random variable with $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) < \infty$, and $a > 0$. Then*

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}. \tag{10}$$

Proof. Note that $\mathbb{P}(|X - \mu| \geq a) = \mathbb{P}((X - \mu)^2 \geq a^2)$ and apply Markov's inequality (Theorem 2.1) to the non-negative random variable $(X - \mu)^2$. □

Next, we provide an example on how Markov's and Chebychev's inequality can be applied on Poisson distributed random variables:

Example 2.3 (Chebychev and Markov for Poisson random variables)**.** *Let $X \sim \mathsf{Poi}(\lambda)$. Markov's inequality implies that*

$$\mathbb{P}(X \geq a) \leq \mathbb{E}[X]/a = \lambda/a, \tag{11}$$

while Chebychev's inequality implies

$$\mathbb{P}(|X - \lambda| > a) \leq \mathrm{Var}(X)/a^2 = \lambda/a^2. \tag{12}$$

In particular, Chebychev's inequality provides a better upper bound than Markov's inequality when a is relatively large with respect to λ.

Exercise 2.1 (A stronger Markov inequality)**.** *Let, X be a random variable with $\mathbb{E}[X^4] < \infty$. Show that*

$$\mathbb{P}(X > a) \leq a^{-4}\mathbb{E}[X^4]. \tag{13}$$

Exercise 2.2 (Exponential Markov inequality, also known as the Chernoff bound)**.** *Let X be a random variable where $\mathbb{E}[\mathrm{e}^X] < \infty$. Show that*

$$\mathbb{P}(X > a) \leq \mathrm{e}^{-a}\mathbb{E}[\mathrm{e}^X]. \tag{14}$$

3 The Law of Large Numbers

In this section we look further into the convergence of i.i.d. random variables. Here we recall that we call a sequence $(X_1, \ldots, X_n)$ i.i.d. when the random variables are independent and identically distributed. Our main result is the following theorem:

Theorem 3.1 (The weak law of large numbers)**.** *Let $X_1, \ldots, X_n$ be an i.i.d. sequence of random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X) < \infty$. Then*

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{d} \mu. \tag{15}$$

Proof. Introduce $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$. By the definition of convergence in distribution, we need to show, for $x < \mu$,

$$F_{\bar{X}_n}(x) \to F_\mu(x) = 0, \tag{16}$$

and, for $x > \mu$,

$$F_{\bar{X}_n}(x) \to F_\mu(x) = 1. \tag{17}$$

As $F_\mu(x)$ is discontinuous in $x = \mu$ we do not need to show convergence there. Next we show convergence in both cases. Note that, for $x < \mu$,

$$F_{\bar{X}_n}(x) - F_\mu(x) = F_{\bar{X}_n}(x) \leq \mathbb{P}(|\bar{X}_n - \mu| > |x - \mu|), \tag{18}$$

and, for $x > \mu$,

$$F_\mu(x) - F_{\bar{X}_n}(x) = 1 - F_{\bar{X}_n}(x) = \mathbb{P}(\bar{X}_n > x) \leq \mathbb{P}(|\bar{X}_n - \mu| > |x - \mu|). \tag{19}$$

Therefore,

$$|F_{\bar{X}_n}(x) - F_\mu(x)| \leq \mathbb{P}(|\bar{X}_n - \mu| > |x - \mu|). \tag{20}$$

It is therefore sufficient to show, for $a > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| > a) \to 0. \tag{21}$$

We can apply Chebychev's inequality (Theorem 2.2):

$$\mathbb{P}(|\bar{X}_n - \mu| > a) = \mathbb{P}(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| > a) \leq \mathrm{Var}(\bar{X}_n)/a^2, \tag{22}$$

where we use that $\mathbb{E}[\bar{X}_n] = \mu$. Further using that $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, we conclude that

$$\mathbb{P}(|\bar{X}_n - \mu| > a) \leq \mathrm{Var}(\bar{X}_n)/a^2 = \sigma^2/(a^2 n) \to 0, \tag{23}$$

as required. □

Exercise 3.1 (Variance of the sample mean)**.** *Suppose that $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ denotes the sample average of the i.i.d. sequence of random variables $(X_i)_{i=1}^n$, with $\mathrm{Var}(X_i) = \sigma^2$. Show that $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$.*

Exercise 3.2 (Convergence of the binomial distribution)**.** *Define $Y_n \sim \mathsf{Bin}(n, p)$. Show that $X_n = Y_n/n \xrightarrow{d} p$.*

Exercise 3.3 (Extensions of the law of large numbers)**.** *Show that the weak law of large numbers in Theorem 3.1 can be extended to the setting where $(X_i)_{i=1}^n$ are independent random variables with $\sup_{i \geq 1} \mathrm{Var}(X_i) < \infty$.*

In Theorem 3.1, we have assumed that $\mathrm{Var}(X_i) < \infty$. This assumption is, however, too strict and Theorem 3.1 still holds if we would only assume $\mathbb{E}[|X_i|] < \infty$. The proof for this theorem is, however, too complicated and outside the scope of this course.

4 The Central Limit Theorem

In this section we discuss the fluctuations occurring in the the law of the large numbers in Theorem 3.1. Note that

$$\mathbb{E}[X_1 + \cdots + X_n] = n\mathbb{E}[X_1], \qquad \text{while} \qquad \text{Var}(X_1 + \cdots + X_n) = n\text{Var}(X_1). \tag{24}$$

It follows that the standard deviation of $X_1 + \cdots + X_n$ is equal to $\sqrt{\sigma^2 n}$, where $\sigma^2 = \text{Var}(X_1)$, while the expectation equals $n\mu$, with $\mu = \mathbb{E}[X_1]$. We conclude that $X_1 + \cdots + X_n$ is concentrated more and more around its mean as n increases. One could argue that this is the reason why statistics works in the first place, as it makes it precise how more data gives rise to higher accuracy.

The central limit theorem (CLT) describes the behaviour of the random variable

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}}. \tag{25}$$

Note that $\mathbb{E}[Z_n] = 0, \text{Var}(Z_n) = 1$, in other words Z_n is *standardized*:

Exercise 4.1 (Z_n is standardized)**.** *show that* $\mathbb{E}[Z_n] = 0, \text{Var}(Z_n) = 1$ *for* Z_n *in* (25)*.*

The following theorem shows that Z_n converges to the standard normal distribution.

Theorem 4.1 (Central limit theorem)**.** *Let* $X_1, \ldots, X_n$ *be a sequence of i.i.d. random variables with* $\mathbb{E}[X_i] = \mu$ *and* $\sigma^2 = \text{Var}(X) \in (0, \infty)$*. Then*

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z, \tag{26}$$

where Z *has a standard normal distribution, i.e.,* $Z \sim \mathcal{N}(0, 1)$*.*

The CLT (Theorem 4.1) is an important theorem for many reasons. In this document we provide three:

Universality. Universality is a concept of physics, which states that many models, that are very diverse on a microscopic level, behave in a similar manner. From a physics perspective, this is an important result. After all, if we model a phenomenon by two models that only differ slightly, then we hope that the conclusion reached within these models are more or less the same. The CLT (Theorem 4.1) is an example of this concept. If we do not model the elements in a large sum correctly, the CLT (Theorem 4.1) states that the sum still behaves according to a normal distribution. The specific distribution of the elements is not relevant. There exist many extensions of the CLT (Theorem 4.1) where, for example, the elements are not from the same distribution, or slightly dependent, where the sum still behaves as a normally distributed random variable. The small details are not that important, we still end with a normal distribution!

Statistical models. In statistics, the normal distribution is a frequently used statistical model. The CLT (Theorem 4.1) shows *why* this statistical model is suitable in many cases. Since many observations can be interpreted as arising from the sum of many small independent effects, the CLT (Theorem 4.1) shows that in such cases the normal distribution is a natural model. Especially when considering that the normal distribution occurs in many other cases as a limiting distribution (see the previous item).

Approximation for probabilities. It can be very hard to *exactly* calculate certain probabilities. For example, the probability that a Poisson random variable with parameter n is greater than or equal to n is equal to

$$\sum_{k \geq n} \mathrm{e}^{-n} \frac{n^k}{k!}, \tag{27}$$

but expressing this sum *exactly*, especially for large values of n, is almost impossible. However, according to the CLT (Theorem 4.1) this probability is approximately $1/2$. Similar problems occur when we look at a binomial random variable, that is, due to the binomial coefficient, hard to express explicitly. Ironically, calculating these probabilities exactly, only becomes *harder* when n is large, but the CLT (Theorem 4.1) gives a more accurate approximation!

Because of the above, the CLT is sometimes called the

fundamental theorem of statistics,

in that it plays a similar role in statistics as the fundamental theorem of calculus does in analysis and the fundamental theorem of algebra in algebra.

In the following chapter we provide a proof of the CLT (Theorem 4.1). This proof is quite complicated, even though it is only based on a couple of analytical steps.

Exercise 4.2 (Assumption of a finite second moment)**.** *Assume X follows a Cauchy distribution. Then the density of X on $\mathbb{R}$ is given by*

$$f_X(x) = \frac{1}{\pi(1+x^2)}. \tag{28}$$

Show that the first moment of X does not exist, and therefore also $\mathbb{E}[X^2] = \infty$. Furthermore, one can show that $X_1 + \cdots + X_n$ has the same distribution as nX (but this computation is beyond the scope of this course, so you may take it as a fact). Show that $X_1 + \cdots + X_n$ does not meet the assumptions of the CLT.

Exercise 4.3 (Lifespan of lamps)**.** *For a certain type of lamps it is known that the mean lifespan is two years. Furthermore, we also know that a lamp does not get worse over time, i.e., if we know that a lamp works at certain moment in time, then average additional time for it to work is again two years. Propose a model, i.e., a suitable probability distribution, for the life span of a lamp.*

Exercise 4.4 (Lifespan of many lamps)**.** *For the lamps that were introduced in Exercise 4.3, give an approximation that the total lifespan of 20 lamps is at least 50 years.*

4.1 Proof of the Central Limit Theorem (Theorem 4.1)

The proof for the CLT (Theorem 4.1) is difficult, but, in our opinion, every mathematics student should have seen this at least once. Applications of the CLT are present in courses on probability theory, statistics and stochastic decision theory. For this reason we emphasise the importance of reading this proof, so that you know why you can use the CLT (Theorem 4.1).

In this section we prove the CLT (Theorem 4.1) with the extra assumption $\mathbb{E}[|X_i|^3] < \infty$. Even though this assumption is not necessary, it makes the proof a little bit easier.

Most proofs of the CLT make use of so-called *generating functions*. These generating functions are outside the scope of this course and a proof involving them is therefore omitted. They are, however, discussed in the follow-up course Stochastic Processes. Proofs that make use of generating functions implicitly use that convergence of generating functions implies convergence in distribution, causing those proofs not to be entirely self-contained. In this chapter we will provide a different self-contained proof that consists of four steps:

Step 1: Standardization and construct what we need to prove. We must show that

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z, \tag{29}$$

where Z is standard normally distributed. Note that

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i, \tag{30}$$

where $Y_i = (X_i - \mu)/\sigma$ denotes the standardized version of X_i. Therefore, $\mathbb{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) = 1$. Thus, it is sufficient to give the proof for a standardized random variable, which we shall assume from now on. Thus, we will prove the Central Limit Theorem (Theorem 4.1) in the case where $\mu = \mathbb{E}[Y_i] = 0, \sigma^2 = \mathrm{Var}(Y_i) = 1$.

In order to show that $Z_n \xrightarrow{d} Z$, we must prove that

$$F_{Z_n}(z) = \mathbb{P}(Z_n \leq z) \to F_Z(z) = \mathbb{P}(Z \leq z), \tag{31}$$

for every value of z of $z \mapsto F_Z(z)$ where F_Z is continuous in z. Since Z is a continuous random variable, the function $z \mapsto F_Z(z)$ is continuous for *every* $z \in \mathbb{R}$, so that we must show (31) for every $z \in \mathbb{R}$.

Step 2: Continuous approximation of the indicator function. We denote

$$F_{Z_n}(z) = \mathbb{E}[\mathbb{1}_{\{Z_n \leq z\}}] = \mathbb{E}[h(Z_n)], \qquad F_Z(z) = \mathbb{E}[\mathbb{1}_{\{Z \leq z\}}] = \mathbb{E}[h(Z)], \tag{32}$$

where $h(x) = \mathbb{1}_{\{x \leq z\}}$. We compare the expected value of a function with argument Z_n to one that has argument Z. The indicator is, however, not a very nice function to use. Indicators are, for example, not continuous. We therefore approximate h by nicer functions.

Set $\varepsilon > 0$. There exist functions $\underline{h}_\varepsilon$ and $\overline{h}_\varepsilon$ such that

(a) $\underline{h}_\varepsilon$ and $\overline{h}_\varepsilon$ are three times continuously differentiable with uniformly bounded third derivatives;

(b) $\underline{h}_\varepsilon(x) = \overline{h}_\varepsilon(x) = h(x)$ if $|x - z| > \varepsilon$;

(c) $\underline{h}_\varepsilon(x) \leq h(x)$ and $\overline{h}_\varepsilon(x) \geq h(x)$ for all $x \in \mathbb{R}$.

For now, we just assume that such functions exist. Later in this document we provide an example on how to construct such functions. When we construct these functions, we have

$$F_{Z_n}(z) = \mathbb{E}[h(Z_n)] \leq \mathbb{E}[\overline{h}_\varepsilon(Z_n)], \qquad F_{Z_n}(z) = \mathbb{E}[h(Z_n)] \geq \mathbb{E}[\underline{h}_\varepsilon(Z_n)]. \tag{33}$$

We then show that for every function f that is three times continuously differentiable with bounded third derivative,

$$\mathbb{E}[f(Z_n)] \to \mathbb{E}[f(Z)]. \tag{34}$$

We elaborate on this in the following step. Because we took $\varepsilon > 0$ arbitrarily, and thus have convergence of $\mathbb{E}[\overline{h}_\varepsilon(Z_n)]$ and $\mathbb{E}[\underline{h}_\varepsilon(Z_n)]$, we can prove the CLT using these functions. We explain this in detail in Step 4. First, and as promised above, we explain how we construct functions $\underline{h}_\varepsilon$ and $\overline{h}_\varepsilon$ as above.

Take

$$u(x) = \mathrm{e}^{-1/x}\mathrm{e}^{-1/(1-x)} \qquad x \in [0,1]. \tag{35}$$

Also define

$$g(x) = \begin{cases} 1, & x < 0; \\ c\displaystyle\int_x^1 u(y)dy, & x \in (0,1); \\ 0, & x > 1, \end{cases} \tag{36}$$

where $c = 1/\int_0^1 u(z)dz$ is such that $g(0) = 1$. The function $x \mapsto g(x)$ maps $\mathbb{R}$ to $[0,1]$, and is non-increasing, starting at 1 and ending at 0. In that sense, it is similar to $x \mapsto \mathbb{1}_{\{x \leq z\}}$. Note further that $x \mapsto g(x)$ is infinitely many times continuously differentiable (Check this!). Then we take

$$\overline{h}_\varepsilon(x) = g((x-z)/\varepsilon), \qquad \underline{h}_\varepsilon(x) = g((x-z+\varepsilon)/\varepsilon). \tag{37}$$

These functions have precisely the properties (a)-(c) that we have assumed before, as you are asked to show in the following exercise:

Exercise 4.5 (Verification of the properties)**.** *Show that the functions defined in* (37) *meet the following properties (where $h(x) = \mathbb{1}_{\{x \leq z\}}$):*
(a) $\underline{h}_\varepsilon$ and $\overline{h}_\varepsilon$ are 3 times continuously differentiable with uniformly bounded third derivatives;
(b) $\underline{h}_\varepsilon(x) = \overline{h}_\varepsilon(x) = h(x)$ if $|x - z| > \varepsilon$;
(c) $\underline{h}_\varepsilon(x) \leq h(x)$ and $\overline{h}_\varepsilon(x) \geq h(x)$ for all $x \in \mathbb{R}$.

Step 3: Convergence for 'nice' functions. Let $x \mapsto f(x)$ be a three times continuously differentiable function with a bounded third derivative, i.e., $\sup_{x \in \mathbb{R}} |f'''(x)| < \infty$. We show that (34) holds for such a function f. Note that $Z \stackrel{d}{=} (W_1 + \cdots + W_n)/\sqrt{n}$, where $W_1, \ldots, W_n$ is an i.i.d. sequence of standard normally distributed random variables.

Exercise 4.6 (Distribution of $(W_1+\cdots+W_n)/\sqrt{n}$)**.** *Show that $(W_1+\cdots+W_n)/\sqrt{n}$ is standard normally distributed if $W_1, \ldots, W_n$ are i.i.d. and standard normally distributed.*

Therefore

$$\mathbb{E}[f(Z)] = \mathbb{E}\Big[f\big((W_1 + \cdots + W_n)/\sqrt{n}\big)\Big]. \tag{38}$$

So we must show that

$$\lim_{n\to\infty} \mathbb{E}\Big[f\big((Y_1 + \cdots + Y_n)/\sqrt{n}\big)\Big] - \mathbb{E}\Big[f\big((W_1 + \cdots + W_n)/\sqrt{n}\big)\Big] = 0. \tag{39}$$

We can show that this equality holds with the use of a Taylor expansion and a telescopic sum. First, we introduce some new notation. Define, for $i \in \{1, \ldots, n-1\}$,

$$U_{i;n} = Y_1 + \cdots + Y_i + W_{i+1} + \cdots + W_n, \tag{40}$$

and $U_{n;n} = Y_1 + \cdots + Y_n$ and $U_{0;n} = W_1 + \cdots + W_n$. It follows that $U_{i;n}$ is the ith interpolation between $U_{n;n} = Y_1 + \cdots + Y_n$ and $U_{0;n} = W_1 + \cdots + W_n$, such that

$$\begin{aligned} &\mathbb{E}\Big[f\big((Y_1 + \cdots + Y_n)/\sqrt{n}\big)\Big] - \mathbb{E}\Big[f\big((W_1 + \cdots + W_n)/\sqrt{n}\big)\Big] \\ &\quad = \mathbb{E}\Big[f\big(U_{n;n}/\sqrt{n}\big)\Big] - \mathbb{E}\Big[f\big(U_{0;n}/\sqrt{n}\big)\Big]. \end{aligned} \tag{41}$$

We can rewrite this as a telescopic sum

$$\begin{aligned} \mathbb{E}[f(Z_n)] - \mathbb{E}[f(Z)] &= \mathbb{E}\Big[f\big(U_{n;n}/\sqrt{n}\big)\Big] - \mathbb{E}\Big[f\big(U_{0;n}/\sqrt{n}\big)\Big] \\ &= \sum_{i=1}^{n} \mathbb{E}\Big[f\big(U_{i;n}/\sqrt{n}\big) - f\big(U_{i-1;n}/\sqrt{n}\big)\Big]. \end{aligned} \tag{42}$$

The factors $U_{i;n}/\sqrt{n}$ and $U_{i-1;n}/\sqrt{n}$ are *precisely* the same, *except* for the ith summand, which is $Y_i/\sqrt{n}$ for $U_{i;n}/\sqrt{n}$ and $W_i/\sqrt{n}$ for $U_{i-1;n}/\sqrt{n}$. Denote the terms that are equal as $S_{i;n} = Y_1 + \cdots + Y_{i-1} + W_{i+1} + \cdots + W_n$. Then

$$\mathbb{E}\Big[f\big(U_{i;n}/\sqrt{n}\big) - f\big(U_{i-1;n}/\sqrt{n}\big)\Big] = \mathbb{E}\Big[f\big((S_{i;n} + Y_i)/\sqrt{n}\big) - f\big((S_{i;n} + W_i)/\sqrt{n}\big)\Big]. \tag{43}$$

Equation (43) is very suggestive. Indeed, the difference in arguments in the two terms becomes increasingly small when n is very large. Thus, (43) begs for a *Taylor expansion*. In turn, this is precisely why we need that f is sufficiently differentiable.

Next we use a Taylor expansion for f, up to third order. Recall that for a Taylor expansion for the function $y \mapsto f(y)$, around x, behaves as

$$f(y) = f(x) + (y-x)f'(x) + \tfrac{1}{2}(y-x)^2 f''(x) + \tfrac{1}{6}(y-x)^3 f'''(t^*), \tag{44}$$

where t^* is a value between x en y. We expand f around $S_{i;n}$ and take $x = S_{i;n}$ and $y = U_{i;n}$, so that

$$f(U_{i;n}/\sqrt{n}) = f(S_{i;n}) + (Y_i/\sqrt{n})f'(S_{i;n}) + \tfrac{1}{2}(Y_i/\sqrt{n})^2 f''(S_{i;n}) + \tfrac{1}{6}(Y_i/\sqrt{n})^3 f'''(t_1^*). \tag{45}$$

This expands the first term in (43). For the second term, we again Taylor expand, now with $x = S_{i;n}$ as before, but $y = U_{i-1;n}$, so that

$$f(U_{i-1;n}/\sqrt{n}) = f(S_{i;n}) + (W_i/\sqrt{n})f'(S_{i;n}) + \tfrac{1}{2}(W_i/\sqrt{n})^2 f''(S_{i;n}) + \tfrac{1}{6}(W_i/\sqrt{n})^3 f'''(t_2^*), \tag{46}$$

for some point t_2^* between $U_{i;n}/\sqrt{n}$ and $S_{i;n}/\sqrt{n}$ and t_2^* between $U_{i-1;n}/\sqrt{n}$ and $S_{i;n}/\sqrt{n}$. This results in the following, quite intimidating, expression:

$$\begin{aligned} &\mathbb{E}[f(Z_n)] - \mathbb{E}[f(Z)] \\ &\quad = \sum_{i=1}^{n} \mathbb{E}\Big[\big(f(S_{i;n}) + (Y_i/\sqrt{n})f'(S_{i;n}) + \tfrac{1}{2}(Y_i/\sqrt{n})^2 f''(S_{i;n}) + \tfrac{1}{6}(Y_i/\sqrt{n})^3 f'''(t_1^*)\big) \\ &\qquad - \big(f(S_{i;n}) + (W_i/\sqrt{n})f'(S_{i;n}) + \tfrac{1}{2}(W_i/\sqrt{n})^2 f''(S_{i;n}) + \tfrac{1}{6}(W_i/\sqrt{n})^3 f'''(t_2^*)\big)\Big] \\ &\quad = \sum_{i=1}^{n} \mathbb{E}\Big[((Y_i - W_i)/\sqrt{n})f'(S_{i;n}) + \tfrac{1}{2n}(Y_i^2 - W_i^2)f''(S_{i;n}) + \tfrac{1}{6}(Y_i/\sqrt{n})^3 f'''(t_1^*) - (W_i/\sqrt{n})^3 f'''(t_2^*)\Big]. \end{aligned} \tag{47}$$

We simplify all of the above defined factors in the sum. Note that Y_i and $S_{i;n}$ are independent, and therefore, Y_i and $f'(S_{i;n})$ are independent as well. Therefore,

$$\mathbb{E}\big[Y_i f'(S_{i;n})\big] = \mathbb{E}[Y_i]\mathbb{E}[f'(S_{i;n})] = 0, \tag{48}$$

because $\mathbb{E}[Y_i] = 0$. By the same argumentation, W_i and $f'(S_{i;n})$ are independent as well and $\mathbb{E}[W_i] = 0$, so that

$$\mathbb{E}\big[W_i f'(S_{i;n})\big] = \mathbb{E}[W_i]\mathbb{E}[f'(S_{i;n})] = 0. \tag{49}$$

Furthermore, due to the independence of Y_i and $f'(S_{i;n})$,

$$\mathbb{E}[Y_i^2 f''(S_{i;n})] = \mathbb{E}[Y_i^2]\mathbb{E}[f''(S_{i;n})] = \mathbb{E}[f''(S_{i;n})], \tag{50}$$

holds, as $\mathbb{E}[Y_i^2] = \mathrm{Var}(Y_i) = 1$. By the same argument,

$$\mathbb{E}[W_i^2 f''(S_{i;n})] = \mathbb{E}[W_i^2]\mathbb{E}[f''(S_{i;n})] = \mathbb{E}[f''(S_{i;n})], \tag{51}$$

holds too, as $\mathbb{E}[W_i^2] = \mathrm{Var}(W_i) = 1$. We conclude that

$$\mathbb{E}\Big[\tfrac{1}{2n}(Y_i^2 - W_i^2) f''(S_{i;n})\Big] = 0, \tag{52}$$

so that

$$\mathbb{E}[f(Z_n)] - \mathbb{E}[f(Z)] = \frac{1}{6n^{3/2}} \sum_{i=1}^{n} \mathbb{E}\Big[Y_i^3 f'''(t_1^*) - W_i^3 f'''(t_2^*)\Big].$$

Clearly, this is far less intimidating! Further, note that the above computations have used all the assumptions on our random variables, in casu that they are independent, that they have mean zero, and that they have variance 1.

Next, we bound this quantity as

$$\begin{aligned}
\Big|\mathbb{E}[f(Z_n)] - \mathbb{E}[f(Z)]\Big| &\leq \frac{1}{6n^{3/2}} \sup_{x\in\mathbb{R}} |f'''(x)| \sum_{i=1}^{n} \mathbb{E}\Big[|Y_i|^3 + |W_i|^3\Big] \\
&= \frac{1}{6n^{3/2}} \sup_{x\in\mathbb{R}} |f'''(x)| n\mathbb{E}\Big[|Y_1|^3 + |W_1|^3\Big] \\
&= \frac{1}{6n^{1/2}} \sup_{x\in\mathbb{R}} |f'''(x)| \mathbb{E}\Big[|Y_1|^3 + |W_1|^3\Big],
\end{aligned} \tag{53}$$

where the second equality again follows from the fact that $\mathbb{E}[|Y_i|^3] = \mathbb{E}[|Y_1|^3]$ and $\mathbb{E}[|W_i|^3] = \mathbb{E}[|W_1|^3] < \infty$.

We see that $\sup_{x\in\mathbb{R}} |f'''(x)|$ is finite by assumption. Furthermore, W_1 is standard normally distributed, so that $\mathbb{E}[|W_1|^3] < \infty$, while we also assumed that $\mathbb{E}[|X_1|^3] < \infty$, so that also $\mathbb{E}[|Y_1|^3] < \infty$.

Define $C = \frac{1}{6} \sup_{x\in\mathbb{R}} |f'''(x)| \mathbb{E}\Big[|Y_1|^3 + |W_1|^3\Big]$, then we have just showed that

$$\Big|\mathbb{E}[f(Z_n)] - \mathbb{E}[f(Z)]\Big| \leq C/\sqrt{n}. \tag{54}$$

As this quantity converges to zero, we have shown that (39) holds for every function f with a uniformly bounded third derivative.

Step 4: Conclusion of the proof. We are now ready to finish the proof. Take $z \in \mathbb{R}$ fixed, define $\varepsilon > 0$, and note

$$\begin{aligned} F_{Z_n}(z) = \mathbb{E}[h(Z_n)] &\leq \mathbb{E}[\overline{h}_\varepsilon(Z_n)] \to \mathbb{E}[\overline{h}_\varepsilon(Z)] \\ &\leq \mathbb{E}[\mathbb{1}_{\{Z \leq z+\varepsilon\}}] = \mathbb{P}(Z \leq z+\varepsilon), \end{aligned} \tag{55}$$

where we now use $\overline{h}_\varepsilon(x) \leq \mathbb{1}_{\{x \leq z+\varepsilon\}}$. In the same way, only now using $\underline{h}_\varepsilon(x) \geq \mathbb{1}_{\{x \leq z-\varepsilon\}}$, we get

$$\begin{aligned} F_{Z_n}(z) = \mathbb{E}[h(Z_n)] &\geq \mathbb{E}[\underline{h}_\varepsilon(Z_n)] \to \mathbb{E}[\underline{h}_\varepsilon(Z)] \\ &\geq \mathbb{E}[\mathbb{1}_{\{Z \leq z-\varepsilon\}}] = \mathbb{P}(Z \leq z-\varepsilon). \end{aligned} \tag{56}$$

Due to the fact that $z \mapsto F_Z(z) = \mathbb{P}(Z \leq z)$ is a continuous function, and $\varepsilon > 0$ was chosen arbitrarily, we see that

$$F_{Z_n}(z) \to F_Z(z) = \mathbb{P}(Z \leq z), \tag{57}$$

as required. □

This proof is difficult, and requires many bigger and smaller steps which all need to be taken in the correct order. What it possibly does not make clear, is *why* the normal distribution appears in the CLT. This is due to the very first step in the proof. There we have used that

$$Z \stackrel{d}{=} (W_1 + \cdots + W_n)/\sqrt{n}, \tag{58}$$

where $W_1, \ldots, W_n$ are i.i.d. and standard normally distributed. Specifically, Z has the same distribution as $(Z_1 + Z_2)/\sqrt{2}$, where Z_1 and Z_2 are from the same distribution as Z. It turns out that this is *only* true for a normal distribution with a mean of zero. This is the reason that the normal distribution must be the limiting distribution in the CLT. Next, we consider some examples of the CLT.

4.2 Examples and Applications of the CLT (Theorem 4.1)

A step-by-step method of working with the CLT. The final exam of 2WS20 always contains a question regarding the CLT (Theorem 4.1). It is, therefore, important to be fully prepared for such questions. These questions are considered to be difficult, so it helps to have a clear approach with steps that one can follow, on how to handle such questions. First, we elaborate on these steps and after that we give some examples where we apply them. The following steps make the application of the CLT fully algorithmic, and following the steps will guarantee that you apply the CLT correctly.

Step 1: Check whether we can apply the CLT (Theorem 4.1). Step 1 consists of an explanation on *why* the CLT is applicable. It is important to remember all the requirements needed for the CLT (Theorem 4.1) to hold. These requirement are that we consider a sum of i.i.d. random variables with a *finite* variance. This needs to be stated explicitly. So the question regards $S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables with finite variance. Give the distribution of X_i, and also denote what the value of n is.

Step 2: Calculate the expected value and the variance of the i.i.d. random variables. In the previous step we have verified that the question at hand regards a certain sum

$S_n = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are i.i.d. random variables with finite variance. In the second step one must find the expected value μ and the variance σ^2 of X_i, as we use these in the CLT (Theorem 4.1). The result of Step 2 is, therefore, a numerical value that represents $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{Var}(X_i)$.

Step 3: Rewrite the requested probability by standardizing the sum S_n. In Step 1 we have verified that we need to determine a probability of the form $\mathbb{P}(S_n \leq s)$, or $\mathbb{P}(S_n \geq s)$ for an appropriate $s \in \mathbb{R}$. In the third step, we rewrite this probability by standardizing S_n. This means that we rewrite, by making use of the values of μ and σ^2 that we have calculated in Step 2,

$$\mathbb{P}(X_1 + \cdots + X_n \leq s) = \mathbb{P}\Big(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{s - n\mu}{\sqrt{n\sigma^2}}\Big). \tag{59}$$

By writing the probability in such a way, we show that we can apply the CLT (Theorem 4.1), which we do in the final step.

Step 4: Apply the CLT (Theorem 4.1). We approximate

$$\mathbb{P}\Big(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{s - n\mu}{\sqrt{n\sigma^2}}\Big) \approx \mathbb{P}\Big(Z \leq \frac{s - n\mu}{\sqrt{n\sigma^2}}\Big) = \Phi(\frac{s - n\mu}{\sqrt{n\sigma^2}}), \tag{60}$$

where $z \mapsto \Phi(z)$ is the distribution function of a standard normal distribution, which we can find in a table. Finally, we must find the value of $\Phi(z)$ for $z = \frac{s-n\mu}{\sqrt{n\sigma^2}}$. Sometimes, we need to rewrite this probability slightly when it is not in a form as covered by the table. For example when $z > 0$, we can make use of the fact

$$\Phi(-z) = 1 - \Phi(z), \tag{61}$$

and

$$\mathbb{P}(Z > z) = 1 - \Phi(z). \tag{62}$$

With these two facts, we can always rewrite $\Phi(\frac{s-n\mu}{\sqrt{n\sigma^2}})$ in a way that can be looked up in the table.

These 4 steps *always* work, and when following them carefully, nothing can go wrong. Next, we apply the 4 steps on a couple of examples:

Guessing in a multiple choice test. A multiple choice exam consists of 20 questions, each with 4 possible answers. A student passes this exam when at least 11 questions are correctly answered. A student contemplates the strategy of not studying, as the probability of passing, when each answer is guessed, is non-zero. Is this a wise strategy?

Step 1: Check if we can apply the CLT (Theorem 4.1). We use the CLT (Theorem 4.1) for a binomial distribution. Let $S_n \sim \mathsf{Bin}(n, p)$. Then we can write S_n as $S_n = I_1 + \cdots + I_n$, where $I_1, \ldots, I_n$ are i.i.d. Bernoulli variables with success probability p. Because $I_1, \ldots, I_n$ are i.i.d. we can apply the CLT. Also, because $I_i \in \{0, 1\}$, these random variables have all moments, and thus in particular a finite variance. In this case $p = 1/4$ en $n = 20$.

Step 2: Calculate the expected value and the variance of the i.i.d. random variables. If $I_i \sim \mathsf{Ber}(p)$, then we know $\mathbb{E}[I_i] = p, \mathrm{Var}(I_i) = p(1-p)$. So in this context, $\mu = \frac{1}{4}$ and $\sigma^2 = (1/4) \cdot (3/4) = \frac{3}{16}$.

Step 3: Rewrite the objective by standardizing the sum S_n. We are interested in $\mathbb{P}(S_{20} \geq 11)$. Thus we rewrite this as

$$\mathbb{P}(S_{20} \geq 11) = \mathbb{P}\Big(\frac{S_{20} - 20 \cdot \mu}{\sqrt{20\sigma^2}} \geq \frac{11 - 5}{\sqrt{20 \cdot (1/4) \cdot (3/4)}}\Big) \approx \mathbb{P}\Big(\frac{S_{20} - 20 \cdot \mu}{\sqrt{20\sigma^2}} \geq 4.647\Big). \quad (63)$$

Step 4: Apply the CLT (Theorem 4.1). We approximate the probability for a passing grade by

$$\mathbb{P}(S_{20} \geq 11) \approx \mathbb{P}\Big(\frac{S_{20} - 20 \cdot \mu}{\sqrt{20\sigma^2}} \geq 4.647\Big) \approx \mathbb{P}(Z \geq 4.647) = 1 - \Phi(4.647). \quad (64)$$

$\Phi(4.647)$ is so close to 1, that it is not even contained in the tables. Hence, this is not a wise strategy.

Exercise 4.7 (Guessing in a multiple choice test with some knowledge)**.** *Now we consider a case where the student studied a little, so that this student is always able to exclude one of four answers. The student chooses an answer of the other 3 at random. Approximate the probability that the student will pass the test, using the CLT.*

The paradox of the deciding minority. In many countries, such as in Holland or in Israel, we see that a small minority can have a big influence on the parliament. In the parliament itself, this is often a small party that helps the coalition to a majority vote. In the next example, we see how a minority can have considerable influence on the outcome of an election. Consider a country with two parties and 1,000,001 voters. We say that a party wins when they get more than half of the total votes. So with 1,000,001 voters, this means that a party needs to have at least 500,001 votes. If every voter chooses one of the two parties with equal probability, both parties have *exactly* the same probability of winning, which is 1/2. After all, the probability of getting at most 499,999 votes equals the probability of getting 500,001 votes (due to the symmetry around 500,000.5 of the binomial distribution with success probability $p = 1/2$), and summed, they must equal 1. So both equal exactly 1/2.

Assume now that one party has a loyal group consisting of 1,000 voters that *always* vote for this party. Give an approximation on the probability that this party wins the election.

Step 1: Check if we can apply the CLT (Theorem 4.1). We use the CLT (Theorem 4.1) for a binomial distribution. Define $S_n \sim \mathsf{Bin}(n, p)$. Then we know that S_n can be written as $S_n = I_1 + \cdots + I_n$, where $I_1, \ldots, I_n$ are i.i.d. Bernoulli random variables with success probability p. As $I_1, \ldots, I_n$ are i.i.d. we can apply the CLT. In this context $n = 999,001$ and $p = 1/2$, and the party with the loyal minority wins if it get 499,001 votes of the remaining total of 999,001 votes.

Step 2: Calculate the expected value and the variance of the i.i.d. random variables. If $I_i \sim \mathsf{Ber}(p)$, then $\mathbb{E}[I_i] = p, \mathrm{Var}(I_i) = p(1-p)$. In this case, we find $\mu = 1/2$ and $\sigma^2 = (1/2) \cdot (1/2) = 1/4$.

Step 3: Rewrite the objective by standardizing the sum S_n. We are interested in $\mathbb{P}(S_{999,001} \geq 499,001)$. We can rewrite this as

$$\begin{aligned}\mathbb{P}(S_{999,001} \geq 499,001) &= \mathbb{P}\Big(\frac{S_{999,001} - 999,001 \cdot \mu}{\sqrt{999,001\sigma^2}} \geq \frac{-499.5}{\sqrt{999,001 \cdot (1/4)}}\Big) \\ &\approx \mathbb{P}\Big(\frac{S_{999,001} - 999,001 \cdot \mu}{\sqrt{999,001\sigma^2}} \geq -0.9994\Big).\end{aligned} \tag{65}$$

Step 4: Apply the CLT (Theorem 4.1). We approximate the probability that the party with the loyal minority wins by

$$\begin{aligned}\mathbb{P}(S_{999,001} \geq 499,001) &\approx \mathbb{P}\Big(\frac{S_{999,001} - 999,001 \cdot \mu}{\sqrt{999,001\sigma^2}} \geq -0.9994\Big) \\ &\approx \mathbb{P}(Z \geq -0.9994) = 1 - \Phi(-0.9994) = \Phi(0.9994) \approx 0.8413.\end{aligned} \tag{66}$$

It follows that a small minority does have a significant influence on the election.

Exercise 4.8 (A larger minority)**.** *Consider now a minority consisting of 2,000. What is the probability that the party with the loyal minority wins?*

An approximation of an integral. We can also use the CLT to approximate integrals. For example, the integral

$$\int_0^n \frac{t^{n-1}}{(n-1)!}\mathrm{e}^{-t}dt \tag{67}$$

can be considered as $\mathbb{P}(S_n \leq n)$, where $S_n = E_1 + \cdots + E_n$ and $(E_i)_{i=1}^n$ are i.i.d. exponentially distributed random variables. So here we can also apply the CLT. As $\mathbb{E}[E_i] = 1$ and $\mathrm{Var}(E_i) = 1$, we see

$$\int_0^n \frac{t^{n-1}}{(n-1)!}\mathrm{e}^{-t}dt = \mathbb{P}(S_n \leq n) = \mathbb{P}\Big(\frac{E_1 + \cdots + E_n - n}{\sqrt{n}} \leq 0\Big) \approx \frac{1}{2}. \tag{68}$$

Exercise 4.9 (Distribution of a sum of exponential random variables)**.** *Show that a sum of n i.i.d. exponentially distributed random variables with parameter λ follows a gamma distribution with parameters n and λ. Conclude that* (68) *holds.*

Exercise 4.10 (The CLT for an integral)**.** *Apply the four steps of the CLT for the above defined example.*

Exercise 4.11 (The CLT for a related integral)**.** *Use the CLT to approximate the integral* $\int_0^{n+\sqrt{n}} \frac{t^{n-1}}{(n-1)!}\mathrm{e}^{-t}dt$.

The CLT for a summation. The sum

$$\sum_{k=0}^{\lambda n} \mathrm{e}^{-\lambda n}\frac{(\lambda n)^k}{k!} \tag{69}$$

can also be interpreted as $\mathbb{P}(S_n \leq \lambda n)$, where now $S_n = X_1 + \cdots + X_n$ and $(X_i)_{i=1}^n$ are i.i.d. Poisson distributed random variables with parameter λ. So the CLT can also be applied here. As $\mathbb{E}[X_i] = \mathrm{Var}(X_i) = \lambda$ we see

$$\sum_{k=0}^{\lambda n} \mathrm{e}^{-\lambda n}\frac{(\lambda n)^k}{k!} = \mathbb{P}(S_n \leq \lambda n) = \mathbb{P}\Big(\frac{X_1 + \cdots + X_n - \lambda n}{\sqrt{\lambda n}} \leq 0\Big) \approx \frac{1}{2}. \tag{70}$$

Exercise 4.12 (The CLT for a summation)**.** *Apply the four steps of the CLT for the above defined example.*

Exercise 4.13 (CLT for a related summation)**.** *Use the CLT to approximate the sum*

$$\sum_{k=0}^{\lambda n-\sqrt{\lambda n}} \mathrm{e}^{-\lambda n}\frac{(\lambda n)^k}{k!}.$$