Stockholms universitet

Edgeworth Expansions for Studying Rate of Convergence towards Normality - with Applications to Variance-Stabilizing Transformations

August Jonasson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2024:4
Matematisk statistik
Maj 2024

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2024:4**
http://www.math.su.se

Edgeworth Expansions for Studying Rate of Convergence towards Normality - with Applications to Variance-Stabilizing Transformations

August Jonasson*

May 2024

Abstract

The aim of this thesis is to present a theoretical framework through which the rate of convergence towards normality can be studied (in the univariate case) - thus providing insight as to what factors to consider when applying the central limit theorem. This is done using Edgeworth expansions of the cumulative distribution function and much focus is placed on cumulants, which they rely upon. We also derive a more general version of the Edgeworth expansion that allows for expectation zero and unit variance to appear only asymptotically, as opposed to the standard expansion that assumes these properties to hold in general. The framework is then applied to variance-stabilizing transformations for two underlying sampling distributions - the Poisson and the exponential. The results tell us that applying the variance-stabilizing transformation in its most simple form does not lead to improved convergence, due to the introduction of a bias and the variance not really being a constant for finite samples. Standardizing the variance-stabilized variable, however, we do see signs of improved convergence for both distributions. In the Poisson case the improvements depend on the product of the rate parameter and the sample size, and in the exponential case the improvements permeate through all tested sample sizes. As such we advise to search for other transformations if the underlying sampling distribution is Poisson, but suggest the standardized variance-stabilized transformation as a viable option if the distribution is exponential.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: aujo8630@gmail.com. Supervisor: Ola Hössjer Johannes Heiny.

Acknowledgements

I would like to express my deepest gratitude towards Ola Hössjer for his invaluable input and unwavering commitment to provide guidance throughout this whole work. I especially want to highlight his crucial role in the parts leading up to the derivation of the generalization of the Edgeworth expansion, equation (2.14), which would have been utterly impossible for me to do on my own. Our weekly meetings, that more often than not stretched way past what was scheduled, are really at the core of this thesis, from beginning to end.

I am also deeply grateful towards Johannes Heiny for his input on the structure of the thesis and his help in identifying errors. The critical feedback he provided lead to significant improvements on clarity, coherence and rigor. His contributions have been truly helpful in improving the quality of this thesis.

Furthermore, I would like to extend my heartfelt gratitude towards Taras Bodnar for his exceptional lectures in statistical inference. He not only sparked my interest for the subject, but also played a pivotal role in inspiring the idea for this thesis.

Thanks to all my friends who brightened each day and kept my sanity in check (for three years). A special thanks to Matilda Galmar for her proof reading, which led to great improvements on clarity throughout the whole thesis.

No chatbots were used in the making of this thesis.

Contents

1. Introduction

1.1. Motivation

The central limit theorem plays an immensely important role in statistics and probability theory. The ability to approximate the distribution of a sum of random variables using a normal distribution allows for far less complex inference when the sample size is (sufficiently) large. The theorem, however, somewhat fails to tell us just how fast the rate of convergence actually is, and what factors we should be aware of before applying it. This thesis aims to present a theoretical framework (to adress such questions) in the form of Edgeworth expansions of the cumulative distribution function.

A very common practice within statistics is to apply some transformation to obtained data in order to make it "more" normal. A natural use for this framework would therefore be to investigate some class of transformations and say if they bring a random variable of interest closer to normality or not. In this thesis, we study variance-stabilizing transformations. Whether these claim to improve inferential tools or not we have not been able to deduce from the introductory literature on statistical inference. It could be that their original purpose (from less compute-ready times) was to reduce complexity. This motivates us to study them further.

1.2. Structure of the thesis

We begin Chapter 2 by briefly presenting some asymptotic results from maximum-likelihood theory - such as the distribution of the maximum-likelihood estimator, the delta method and the variance-stabilizing transformation. We then introduce the Edgeworth expansion in its base form, give a pretty detailed presentation of cumulants - of which the Edgeworth expansion rely upon - before returning to the Edgeworth expansion and presenting it in a more general form. We proceed with a brief introduction of the natural exponential family and why it forms a nice class of examples to study through this framework. At the very end of the chapter we give a brief run-through of why improved convergence implies improved inference, using confidence intervals and coverage probability.

In Chapter 3 we apply the variance-stabilizing transformation for two underlying sample distributions - the Poisson and the exponential - and use our framework to study the convergence towards normality. The results are presented in tables and some brief interpretations of them are given. We then suggest an improved version of the variance-stabilizing transformation, apply it, and present the new results.

Lastly, we conclude the thesis in Chapter 4 by discussing cumulants, the usage of the Edgeworth expansion in a more practical sense and how it, while being quite cumbersome to work with, might be challenging to replace with simpler, more direct methods of studying convergence towards normality. We present some of the possible shortcomings of the Edgeworth expansion and the closely related Berry-Esseen bound as an alternative way of studying the central limit theorem. We also discuss variance-stabilizing transformations, their drawbacks and possible improvements on them. For further work we suggest, in part, simulation as a means of studying convergence towards normality, but also to continue with the framework presented, but applied to a wider range of transformations.

2. Theoretical Framework

In this chapter we present all of the required theory in order to study finite sample convergence towards normality for some (univariate) random variable whose asymptotic distribution is normal. This will allow us to study the effects of the variance-stabilizing transformation, which will be our starting point.

2.1. MLE asymptotics and the delta method

In order to define the variance-stabilizing transformation, we first have to define the delta method, which in turn relies on the asymptotic distribution of the maximum likelihood estimator.

From here on out, we will assume that the Fisher regularity conditions hold and that $X_{1:n} = (X_1 \dots X_n)$ is a random sample of independent and identically distributed random variables. Furthermore, assume that $\hat{\theta} = \hat{\theta}(\overline{X}_{1:n})$ is the maximum likelihood estimator of θ, which in turn yields that $\hat{\theta}$ is consistent for θ (Held and Bové 2020). It follows that the asymptotic distribution of the maximum likelihood estimator satisfies

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, J^{-1}(\theta)), \quad n \to \infty, \tag{2.1}$$

where $\xrightarrow{D}$ denotes convergence in distribution (Gut 2009, p. 147), and $J(\theta)$ denotes the expected Fisher information (Held and Bové 2020, p. 81) with respect to the distribution of the sample variable X_i.

The delta method now states that, if $\hat{\theta}$ is a statistic such that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \sigma^2(\theta)), \quad \sigma(\theta) > 0, \quad n \to \infty,$$

and g is a once differentiable function $\mathbb{R} \to \mathbb{R}$ with $g'(\theta) \neq 0$ for all θ. Then

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{D} N(0, [g'(\theta)]^2\sigma^2(\theta)) \tag{2.2}$$

(DasGupta 2008, p. 40).

Remark 2.1. Variance-stabilizing transformations only work in the univariate case, hence, we are restricted to only studying univariate transformations. It is, however, worth mentioning that there exists an equivalent definition of the delta method for the multivariate case.

2.2. Variance-stabilizing transformations (VST)

The main idea behind VST is to make the estimator's variance asymptotically independent of the unknown parameter, thereby reducing the amount of unknowns when creating confidence intervals and testing hypotheses. Whether the coverage probability is actually improved or not remains to be seen.

This is where the delta method, (2.2), comes into play. By solving the equation

$$c^2 = [g'(\theta)]^2\sigma^2(\theta),$$

for some constant c, we can find such a transformation g. This in turn gives us an explicit formula

$$g(\theta) = c\int \frac{1}{\sigma(\theta)}\,d\theta, \tag{2.3}$$

in order to find a variance-stabilizing transformation (DasGupta 2008, p. 50). Here, the integral sign denotes a primitive function, and the constant c can be chosen in whichever way simplifies the transformation the most.

Remark 2.2. The variance-stabilizing transformation is a monotone function such that if we make inferences about $g(\theta)$ it is always possible to take the inverse of g in order to make inferences about θ (see DasGupta 2008, p. 50).

Example 2.1. Let $X_{1:n}$ be an iid sample of Poisson variables with rate parameter λ. The maximum likelihood estimator of λ is

$$\hat{\lambda} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and the asymptotic distribution evaluates to

$$\sqrt{n}(\overline{X}_n - \lambda) \xrightarrow{D} N(0,\lambda), \quad n \to \infty.$$

In order to find the VST we use (2.3) and solve

$$g(\lambda) = c\int \frac{1}{\sqrt{\lambda}} d\lambda,$$

thus yielding $g(\lambda) = 2c\sqrt{\lambda}$. Choosing $c = 1/2$ we get the VST $g(\lambda) = \sqrt{\lambda}$. We can now use the delta method (2.2) in order to retrieve the new asymptotic distribution

$$\sqrt{n}\left(\sqrt{\overline{X}_n} - \sqrt{\lambda}\right) \xrightarrow{D} N(0, 1/4),$$

which we can see is now independent of the unknown parameter λ.

2.3. The Edgeworth expansion (EE)

In this section we will give a brief overview of the Edgeworth expansion and state some useful facts. We will not attempt to derive it in its entirety yet, but will instead focus on breaking it down and getting acquainted with its components (for an in depth theoretical derivation, see Barndorff-Nielsen and Cox 1989, ch. 4).

Let

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}, \tag{2.4}$$

where $\mu = E[X_i]$, $\sigma = \sqrt{\mathrm{Var}(X_i)}$ and $\overline{X}_n$ is the sample mean.

Now, the two-term Edgeworth expansion of the cumulative distribution function (CDF) of Z_n can be expressed in the following way (Barndorff-Nielsen and Cox 1989, p. 90):

$$F_{Z_n}(x) = \Phi(x) - \phi(x)\left(\frac{\rho_3 H_2(x)}{6\sqrt{n}} + \frac{\rho_4 H_3(x)}{24n} + \frac{\rho_3^2 H_5(x)}{72n}\right) + \mathcal{O}(n^{-3/2}), \qquad (2.5)$$

where Φ and ϕ denote the standard normal CDF and standard normal density function respectively, $\rho_r = \kappa_r/\sigma^r$, $r = 3, 4, \ldots$, denote the standardized cumulants - κ_r the regular cumulants - of order r of the underlying sample distribution (more on cumulants in the next section), and $\mathcal{O}(a_n)/a_n$ is bounded as $n \to \infty$ for any a_n. The $H_i(x)$ factors originate from a set of orthogonal polynomials called the Hermite polynomials and are defined as (Barndorff-Nielsen and Cox 1989, p. 18)

$$\phi(x)H_r(x) = (-1)^r \frac{d^r}{dx^r}\phi(x), \qquad r = 0, 1, 2 \ldots. \qquad (2.6)$$

Since these polynomials appear in the same form, in each Edgeworth expansion, independently of the underlying distribution, we will henceforth just consider them as a given. However, when performing numerical evaluations of the EE - as we shall do at a later stage - we will have to be a bit mindful of these, since they evaluate to zero for some values on x, and therefore lead to entire terms disappearing from (2.5). The first eight Hermite polynomials are

$$\begin{aligned} H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x, \\ H_4(x) = x^4 - 6x^2 + 3, \quad H_5(x) = x^5 - 10x^3 + 15x, \\ H_6(x) = x^6 - 15x^4 + 45x^2 - 15, \quad H_7(x) = x^7 - 21x^5 + 105x^3 - 105x. \end{aligned} \qquad (2.7)$$

Proposition 2.1. If the Edgeworth expansion exists for a random variable X, then there also exists Edgeworth expansions for transformations of X (Barndorff-Nielsen and Cox 1989, p. 121-122).

The result from Proposition 2.1 is important in the sense that as long as the Edgeworth expansion exists for Z_n (2.4), then it will also exist for transformations of Z_n - in our case variance-stabilizing transformations. In this thesis, we will only consider variables whose Edgeworth expansions exist.

2.4. Cumulants

Looking at the Edgeworth expansion, as presented in (2.5), we can see that, for fixed x, it really only depends on the sample size n and the standardized cumulants ρ_r. In order to define cumulants, we start by defining the moment-generating function.

Definition 2.1. The **moment-generating function (MGF)** of a random variable X is defined as

$$M_X(t) = E[e^{tX}],$$

provided there is some $h > 0$ such that this expectation exists for $|t| < h$ (Gut 2009, p. 63).

Definition 2.2. The **moments** (sometimes referred to as raw moments) of a random variable X whose moment-generating function exists are defined as

$$\mu_r = E[X^r] = M_X^{(r)}(0), \qquad r = 1, 2, \ldots$$

(Gut 2009, p. 64).

Having the MGF and the moments well defined we can now move on to defining the cumulants.

Definition 2.3. The **cumulant-generating function (CGF)** of a random variable X is defined as

$$K_X(t) = \log(M_X(t)),$$

i.e., the logarithm of the moment-generating function (Barndorff-Nielsen and Cox 1989, p. 6).

Remark 2.3. It is also possible (and even more general) to define the cumulant-generating function in terms of the characteristic function $\varphi(t) = E[e^{itX}]$. However, we will only consider distributions whose moments are well defined, hence we settle for the definition using the moment-generating function.

Definition 2.4. The **cumulants** of a random variable X are defined as the coefficients κ_r $(r = 1, 2, \dots)$ in the series expansion

$$K_X(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!},$$

of the cumulant generating function, or, equivalently

$$\kappa_r = K_X^{(r)}(0),$$

i.e, the r:th derivative of the cumulant-generating function, evaluated at zero.

Remark 2.4. Sometimes, when we are working with variables that represent transformations of entire samples (e.g. Z_n in (2.4)), it might be more productive to study the cumulants of the underlying sample distribution. In these cases we will adopt the notation $\kappa_r = \kappa_r(X_i)$, or $\kappa_r = \kappa_r(Z_n)$ to specify which cumulants we are referring to. This will be especially useful when such transformations are affine and involve sums of random variables, as will become evident later on.

Taking the derivative of the CGF several times, or identifying the coefficients in the series expansion may, however, turn out to be rather cumbersome. In these cases we can enjoy the fact that the cumulants can be expressed as functions of the central moments. The first four cumulants are given as

$$\begin{aligned}
\kappa_1 &= E[X] \\
\kappa_2 &= E[(X - E[X])^2] \quad (= \mathrm{Var}(X)) \\
\kappa_3 &= E[(X - E[X])^3] \\
\kappa_4 &= E[(X - E[X])^4] - 3(E[(X - E[X])^2])^2,
\end{aligned} \tag{2.8}$$

which in turn can be further reduced to the raw moments

$$\begin{aligned}
\kappa_1 &= \mu_1 \\
\kappa_2 &= \mu_2 - \mu_1^2 \\
\kappa_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \\
\kappa_4 &= \mu_4 - 4\mu_1\mu_3 + 12\mu_1^2\mu_2 - 3\mu_2^2 - 6\mu_1^4,
\end{aligned} \tag{2.9}$$

for $\mu_r = E[X^r]$. We settle for the first four cumulants here because this is the highest order we will use in our Edgeworth expansions.

Remark 2.5. As is evident from (2.8) the first cumulant is the expectation and the second cumulant is the variance. The third cumulant is the non-standardized skewness of the distribution and "improvements" on this might be referred to as improvements on the skewness. The fourth cumulant is the difference between the fourth centralized moment - the non-standardized kurtosis[1] - and three times the squared variance. For simplicity, we will henceforth refer to this as the kurtosis, short for kurtosis-correction factor.

When deriving the cumulants, we have some additional properties that might come in handy - namely

$$\begin{aligned}\kappa_r(X+c) &= \kappa_r(X)+c, \quad r=1 && \\ \kappa_r(X+c) &= \kappa_r(X), \quad r>1 && \textbf{[translational invariance]} \\ \kappa_r(cX) &= c^r\kappa_r(X) && \textbf{[homogeneity of order r]} \\ \kappa_r(X_1+\ldots+X_n) &= \kappa_r(X_1)+\ldots+\kappa_r(X_n) && \textbf{[cumulative]}.\end{aligned} \tag{2.10}$$

Example 2.2. Let $X_{1:n}$ be an iid sample of some distribution whose MGF exists. Now consider the CGF of the standardized sample variable Z_n (see (2.4)). After some calculations, using the properties of the moment-generating function (Gut 2009, ch. 3.3) we can see that

$$K_{Z_n}(t) = -\frac{\sqrt{n}\mu t}{\sigma} + nK_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right).$$

Hence, the CGF of the underlying sample variable appears in the expression of the CGF of Z_n, thus making it more easily derivable. If we now want the higher-order cumulants of Z_n, using Definition 2.4, the first term disappears after the second differentiation and we get

$$\begin{aligned}\kappa_r(Z_n) &= \frac{d^r}{dt^r}\left[nK_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]\Big|_{t=0} \\ &= n\kappa_r\left(\frac{X_i}{\sigma\sqrt{n}}\right) \\ &\overset{(2.10)}{=} \frac{n}{(\sigma\sqrt{n})^r}\kappa_r(X_i), \qquad r=2,3,\ldots.\end{aligned}$$

That is, the cumulants of Z_n can be expressed as some sample dependent factor multiplied with the cumulant of the underlying sample distribution. This is in part due to the cumulative property of (2.10) and will in general not hold after a variance stabilizing transformation.

The above example highlights an issue with the Edgeworth expansion, as given in equation (2.5). The coefficients $\rho_r = \kappa_r(X_i)/\sigma^r$ take advantage of the cumulative property from (2.10) in the sense that that the cumulants of Z_n are easily relatable to the cumulants of its underlying sample distribution. When this cumulative property does not hold (as will become evident when we explore some examples of VST), the expansion in (2.5) does not not hold either. We are therefore in need of a generalization of the Edgeworth expansion that is directly dependent on the cumulants of the sample variable whose CDF we want to expand.

[1] 2024-05-23: https://en.wikipedia.org/wiki/Kurtosis

2.5. Generalization of the Edgeworth expansion

Another restriction of the Edgeworth expansion from (2.5) is that it hinges on the fact that $E[Z_n] = 0$ and $\text{Var}(Z_n) = 1$, i.e., that our sample variable has standard expectation and variance. This in turn yields an Edgeworth expansion that is free from the first and second cumulants (the expectation and the variance). However, when working with variance-stabilizing transformations we are using the delta method (2.2), which relies solely on asymptotic results - for the distribution, as well as for the expectation and variance. It would therefore be beneficial to have an expansion where the first and second cumulants are included as well, since these deviations from standard expectation and variance might not be negligible.

Let $T_n \xrightarrow{D} N(0,1)$ differ from our previous Z_n in the sense that it is not necessarily truly standardized, only asymptotically. We now want to derive an expansion for F_{T_n} using its cumulants. Using Definition 2.4 we start from the cumulant-generating function

$$K_{T_n}(t) = \kappa_1 t + \kappa_2 \frac{t^2}{2} + \kappa_3 \frac{t^3}{6} + \kappa_4 \frac{t^4}{24} + \ldots,$$

including only the first four cumulants here as well. Adding and subtracting $t^2/2$ and exponentiation gives us the moment-generating function

$$M_{T_n}(t) = \exp\left\{\frac{t^2}{2}\right\} \exp\left\{\kappa_1 t + (\kappa_2 - 1)\frac{t^2}{2} + \kappa_3 \frac{t^3}{6} + \kappa_4 \frac{t^4}{24} + \ldots\right\}.$$

The three-term Maclaurin expansion $e^x = 1 + x + x^2/2$ where $x = \kappa_1 t + (\kappa_2 - 1)t^2/2 + \kappa_3 t^3/6 + \kappa_4 t^4/24$ yields an expression

$$M_{T_n}(t) = \exp\left\{\frac{t^2}{2}\right\} \left(1 + A_1 t + \ldots + A_8 t^8 + \ldots\right),$$

where the coefficients $A_1, \ldots, A_8$ are functions of the cumulants. After some tedious calculations (see Appendix A.1) we retrieve

$$\begin{aligned} A_1 &= \kappa_1, \quad A_2 = \frac{1}{2}(\kappa_1^2 + \kappa_2 - 1), \quad A_3 = \frac{\kappa_3}{6} + \frac{(\kappa_2 - 1)\kappa_1}{2}, \\ A_4 &= \frac{(\kappa_2 - 1)^2}{8} + \frac{\kappa_1\kappa_3}{6} + \frac{\kappa_4}{24}, \quad A_5 = \frac{\kappa_1\kappa_4}{24} + \frac{(\kappa_2 - 1)\kappa_3}{12}, \\ A_6 &= \frac{\kappa_3^2}{72} + \frac{(\kappa_2 - 1)\kappa_4}{48}, \quad A_7 = \frac{\kappa_3\kappa_4}{144}, \quad A_8 = \frac{\kappa_4^2}{1152}. \end{aligned} \tag{2.11}$$

Remark 2.6. Notice how A_1, A_2 and A_5 evaluate to zero when $\kappa_1 = 0$ and $\kappa_2 = 1$. Notice also how the remaining coefficients simplify significantly under this condition. We get $A_3 = \kappa_3/6$, $A_4 = \kappa_4/24$, $A_6 = \kappa_3^2/72$, while A_7 and A_8 remain unchanged.

Now assuming $\kappa_r - \mathbb{I}_{r=2} = \mathcal{O}(n^{-1/2})$ for $r = 1, 2, 3, 4$ it follows that $(\kappa_r - \mathbb{I}_{r=2})(\kappa_m - \mathbb{I}_{m=2}) = \mathcal{O}(n^{-1})$ for $r, m = 1, 2, 3, 4$, where $\mathbb{I}_{i=j}$ is the indicator variable that takes on the value 1 when $i = j$ and 0 otherwise. Using the three-term Maclaurin expansion we are getting all of the cumulant-products of this form, and we can write

$$M_{T_n}(t) = \exp\left\{\frac{t^2}{2}\right\} (1 + A_1 t + \ldots + A_8 t^8 + \mathcal{O}(n^{-3/2})). \tag{2.12}$$

The next step is to invert (2.12) in order to retrieve the density function. To do this, we use the fact that

$$\int e^{tx}\phi(x)H_r(x)dx = t^r \exp\left\{\frac{t^2}{2}\right\}, \qquad t \in \mathbb{R}$$

(Barndorff-Nielsen and Cox 1989, p. 91) where $H_r(x)$ is the Hermite polynomial of degree r as in (2.5) and $\phi(x)$ is the standard normal density. Now, using the definition of Hermite polynomials as in (2.6), one can retrieve an expansion for the density

$$f_{T_n}(x) = \phi(x)\left(1 + A_1H_1(x) + \dots A_8H_8(x)\right) + \mathcal{O}(n^{-3/2}). \tag{2.13}$$

Using (2.6) once again, one can retrieve the cumulative distribution function as

$$F_{T_n}(x) = \Phi(x) - \phi(x)\left(A_1H_0(x) + \ldots + A_8H_7(x)\right) + \mathcal{O}(n^{-3/2}) \tag{2.14}$$

(see Barndorff-Nielsen and Cox 1989, p. 91 for similar progression).

It would now be appropriate to compare (2.14) to (2.5) in order to see if our generalization works. An initial test we might want to do is to just write out (2.14) under the assumption that $\kappa_1 = 0$ and $\kappa_2 = 1$ and see if it looks similar to (2.5). Using the coefficients from Remark 2.6, we get

$$F_{T_n}(x) = \Phi(x) - \phi(x)\left(\frac{\kappa_3H_2(x)}{6} + \frac{\kappa_4H_3(x)}{24} + \frac{\kappa_3^2H_5(x)}{72} + \frac{\kappa_3\kappa_4H_6(x)}{144} + \frac{\kappa_4^2H_7(x)}{1152}\right) + \mathcal{O}(n^{-3/2}).$$

It looks fairly similar, apart from the last two terms in the parenthesis, which do not appear in (2.5). This motivates us to see how the two expansions actually perform in relation to one another, and the standard normal CDF. Since (2.5) only allows for expansion of standardized variables, we will have to choose a T_n with expectation zero and variance one.

For example, let

$$T_n \overset{(2.4)}{=} Z_n.$$

We know that $T_n \xrightarrow{D} N(0,1)$ and that $E[T_n] = 0$, $\mathrm{Var}(T_n) = 1$. Hence $\kappa_1(T_n) = 0$ and $\kappa_2(T_n) = 1$. The idea is now to evaluate the two different expansions (2.5) and (2.14) for some distribution F and some values x and compare them to the standard normal CDF to see if they behave similarly (and if they approximate the standard normal distribution well). Let F^B denote expansion (2.5) (B for Barndorff) and let F^G denote expansion (2.14) (G for generalized). We then get the two Edgeworth expansions

$$\begin{aligned} F^B_{T_n}(x) = \Phi(x) - \phi(x)&\left(\frac{\kappa_3(X_i)H_2(x)}{6\sigma^3\sqrt{n}} + \frac{\kappa_4(X_i)H_3(x)}{24\sigma^4 n} + \frac{\kappa_3(X_i)^2H_5(x)}{72\sigma^6 n}\right) \\ &+ \mathcal{O}(n^{-3/2}), \end{aligned} \tag{2.15}$$

$$\begin{aligned} F^G_{T_n}(x) = \Phi(x) - \phi(x)&\left(\frac{\kappa_3(T_n)H_2(x)}{6} + \frac{\kappa_4(T_n)H_3(x)}{24} + \frac{\kappa_3(T_n)^2H_5(x)}{72}\right. \\ &\left. + \frac{\kappa_3(T_n)\kappa_4(T_n)H_6(x)}{144} + \frac{\kappa_4^2(T_n)H_7(x)}{1152}\right) + \mathcal{O}(n^{-3/2}), \end{aligned} \tag{2.16}$$

to evaluate numerically, where we have used the fact that $\rho_r = \kappa_r/\sigma^r$ for $r = 3, 4, \ldots$ in (2.15).

Using $X_i \sim Po(\lambda)$ we get the following results for some (equidistributed) values on x, while keeping $\lambda = 5$ and $n = 50$ constant:

Table 2.1.: Comparison of Edgeworth expansions for standardized T_n when sampling distribution is $Po(\lambda = 5)$ and $n = 50$ against the standard normal CDF.

x	$F^B_{T_n}(x)$	$\Phi(x)$	$F^G_{T_n}(x)$
-2.00	0.02101	0.02275	0.02101
-1.65	0.04756	0.04947	0.04756
-1.30	0.09552	0.09680	0.09551
-0.95	0.17132	0.17106	0.17132
-0.60	0.27654	0.27425	0.27654
-0.25	0.40514	0.40129	0.40515
0.10	0.54396	0.53983	0.54397
0.45	0.67663	0.67364	0.67664
0.80	0.78922	0.78814	0.78922
1.15	0.87425	0.87493	0.87424
1.50	0.93154	0.93319	0.93153
1.85	0.96605	0.96784	0.96605

From Table 2.1 we can see that, out of the twelve rows in total, in six of them $F^B_{T_n}$ and $F^G_{T_n}$ are equal, and in the remaining six they only differ from each other by an order of 10^{-5}. Both of them have a maximum difference from Φ of order 10^{-3} - they seem to be approximating the normal distribution rather well. The important thing to take away from this result is that $F^G_{T_n}$ appears to work just as well as $F^B_{T_n}$ when T_n is standardized. This will make comparisons between standardized and transformed variables a bit easier, since we can use the same Edgeworth expansion, (2.14), for both of them.

2.6. The natural exponential family

Since our main interest somewhat lies in cumulants, it would be nice to choose a family whose cumulant-generating functions are easily expressible. One such family is the natural exponential family.

Definition 2.5. The **natural exponential family (NEF)** (univariate case) consists of all distributions whose densities can be expressed as

$$f(x;\theta) = h(x)\exp\{x\eta(\theta) - A(\eta(\theta))\},$$

where $\eta(\theta)$ represents the "natural parameter", a parametrization of the distribution parameter θ.

Remark 2.7. When talking about the natural parameter, we might simply refer to it as η, even though it is technically a function of the distribution parameter θ.

Example 2.3. For $X \sim Bernoulli(\pi)$:

$$\begin{aligned} f(x;\pi) &= \pi^x(1-\pi)^{1-x} \\ &= \exp\left\{x\ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)\right\} \end{aligned}$$

We can now identify $\eta(\pi) = \ln(\pi/(1-\pi))$, and $A(\eta(\pi)) = -\ln(1-\pi)$, from which we get $A(\eta) = \ln(1+e^\eta)$.

2.6.1. Moments and cumulants of NEF

The moments and cumulants of the NEF distributions are rather concisely expressed, and this is why they form a nice class of examples to look at. Starting from the moment-generating function: let X belong to the natural exponential family. We have that

$$M_X(t) = E[e^{tX}] = \int_X h(x)\exp\{xt + x\eta - A(\eta)\}dx.$$

By adding and subtracting $A(\eta + t)$ to the exponent, the right-hand side can be rewritten as

$$\exp\{A(\eta + t) - A(\eta)\}\int_X h(x)\exp\{x(\eta + t) - A(\eta + t)\}dx,$$

from which we can use the normalizing condition of densities and retrieve

$$M_X(t) = \exp\{A(\eta + t) - A(\eta)\}. \tag{2.17}$$

From (2.17) follows the cumulant-generating function

$$K_X(t) = A(\eta + t) - A(\eta), \tag{2.18}$$

Furthermore, since the r:th cumulant refers to the r:th derivative of the CGF, evaluated at $t = 0$, we can also express the cumulants themselves rather explicitly. Upon taking the derivative, the second term of (2.18) vanishes, and we get

$$\kappa_r(X) = \frac{d^r}{dt^r}\Big[A(\eta + t)\Big]\Big|_{t=0}, \quad (r = 1, 2, \dots). \tag{2.19}$$

This will be useful in the following section.

2.6.2. Variance-stabilizing transformations for NEF

Recalling that the second cumulant is equal to the variance, we can formulate a direct method of retrieving the VST of a random variable X belonging to the natural exponential family. Using this and (2.19) we get that

$$\mathrm{Var}(X) = \frac{d^2}{dt^2}\Big[A(\eta + t)\Big]\Big|_{t=0},$$

from which it now follows, using (2.3), that the VST can be expressed as

$$g(\theta) = k\int\left(\frac{d^2}{dt^2}\Big[A(\eta(\theta) + t)\Big]\Big|_{t=0}\right)^{-1/2} d\theta, \tag{2.20}$$

for some constant k.

In summary, if we have the density function expressed in terms of its natural exponential parametrization, we can derive the expression for its CGF, and in turn, its VST rather directly.

2.7. Summarizing the main theory

We have now presented the main theory of this thesis. We have a generalized version of the Edgeworth expansion (2.14) that allows for all three distribution properties (normality, expectation zero and unit variance) to only appear asymptotically, while still retaining the advantages of having a truly standardized variable. This expansion is also more general in the sense that it is a function of the cumulants of the variable whose cumulative distribution function we want to expand (2.11), instead of relying on linearity and cumulativity (2.10). We have presented three general ways of acquiring the cumulants: as coefficients in a series expansion of the cumulant-generating function, as derivatives of the cumulant-generating function - Definition 2.4 - and as functions of the raw moments (2.9).

So what do we do now? The idea is to try the Edgeworth expansion out as a tool for examining the rate of convergence towards normality. For this, we have specifically chosen to study variance stabilizing transformations (2.3), since these take some standardized variable T_n, and through some transformation $g(T_n)$, make its variance asymptotically independent of its unknown parameter. Through the delta method (2.2) this comes at the price of loosing the standardization - therefore forming a nice basis of examples to look at through the lens of this framework. To make things a bit easier, we have also chosen a specific family of distributions - the natural exponential family - whose cumulant-generating functions seem to be rather easily retrieved (2.18).

2.8. How we apply the theory

Looking at at any of the Edgeworth expansions we have presented so far, we can see that they all start with the term $\Phi(x)$, the cumulative distribution function of a standard normal distribution. If the variable whose CDF we want to expand is truly standard normal, i.e., such that $\kappa_1 = 0$, $\kappa_2 = 1$, $\kappa_3 = \kappa_4 = \kappa_5 = \ldots = 0$ (the normal distribution is the only distribution with this property), all that comes after this first term would evaluate to zero - thus leaving a perfect equality between the left- and right-hand sides. This, somewhat loosely, implies that if whatever comes after this first term is close to zero, the distribution is close to normal. Hence, one approach would be to examine

$$\sup_{x\in I}\left|\sum_{j=1}^{8} A_j H_{j-1}(x)\right| \tag{2.21}$$

both pre- and post-VST, for some interval I that accounts for most of the probability mass of a normal distribution. This would, however, be a bit conservative in the sense that the supremum does not really tell us the whole story. It only tells us where the deviations from normality are the largest. It could very well be the case that the VST is closer to normal in some sub-interval of I while further away in another. In order to loosen this up a bit, we could therefore make sure to put several different sub-intervals of I through (2.21). This would then paint a broader picture of how the VST actually affects the normality.

For the sub-intervals of I we decide on five (approximate) percentiles of the standard normal distribution - the 5th, 40th, 60th and 95th. Setting $I = [-5, 5]$, which accounts for $1 - c \cdot 10^{-7}$ of the probability mass of the standard normal distribution for some constant c, we then have the sub-intervals

$$\begin{aligned} I_1 &= [-5, -1.64] \\ I_2 &= [-1.64, -0.25] \\ I_3 &= [-0.25, 0.25] \\ I_4 &= [0.25, 1.64] \\ I_5 &= [1.64, 5] \end{aligned} \tag{2.22}$$

to examine. Using all of this, we are now ready to jump into some examples. Before we do that, however, we will briefly try to concretize why improved convergence towards normality implies improved inference.

2.9. Implications of improved convergence

So far we have only assumed that improved convergence towards normality of the cumulative distribution function of some asymptotically normal variable T_n implies improvements for inferential tools. In this last section of the theory we will try to make these improvements a bit more concrete by showing how improved convergence affects the difference between nominal and actual coverage.

Let $T_n = T_n(\theta) \xrightarrow{D} N(0,1)$ be our (transformed) sample statistic - a function of our sample and the unknown parameter θ - with cumulative distribution function F_{T_n}. Also let Φ denote the CDF of a standard normal distribution. A two-sided $100\% \cdot (1-\alpha)$ confidence interval for θ would then, under the normality assumption, be

$$I_\theta = \{\theta : x_1 \leq T_n(\theta) \leq x_2\}, \tag{2.23}$$

where $x_1 = \Phi^{-1}(\alpha/2)$ and $x_2 = \Phi^{-1}(1-\alpha/2)$ are, respectively, the $\alpha/2$ and $1-\alpha/2$-quantiles of the standard normal distribution. Now, we want to look at the actual coverage probability of I_θ and show that the difference between the actual coverage and the nominal coverage decreases as F_{T_n} approaches Φ. The true coverage probability is defined as

$$P_n(\theta) = \mathbb{P}(\theta \in I_\theta) = \mathbb{P}(x_1 \leq T_n \leq x_2) = F_{T_n}(x_2) - F_{T_n}(x_1). \tag{2.24}$$

Using the fact that

$$(1-\alpha) - (\Phi(x_2) - \Phi(x_1)) = (1-\alpha) - \left(1 - \frac{\alpha}{2} - \frac{\alpha}{2}\right) = 0,$$

we can rewrite (2.24) to

$$P_n(\theta) = (1-\alpha) + (F_{T_n}(x_2) - \Phi(x_2)) - (F_{T_n}(x_1) - \Phi(x_1)),$$

which is equivalent to

$$P_n(\theta) - (1-\alpha) = (F_{T_n}(x_2) - \Phi(x_2)) - (F_{T_n}(x_1) - \Phi(x_1)).$$

Now applying the triangle-inequality, this yields

$$\begin{aligned} |P_n(\theta) - (1-\alpha)| &\leq |F_{T_n}(x_2) - \Phi(x_2)| + |F_{T_n}(x_1) - \Phi(x_1)| \\ &\leq 2 \sup_{x \in \mathbb{R}} \big| F_{T_n}(x) - \Phi(x) \big|, \end{aligned} \tag{2.25}$$

where, in the last step, we have used the fact that the sum of two distances has to be less or equal to twice the maximal possible distance.

Interpreting (2.25) we can see that the left-hand side is the absolute difference between the nominal and actual coverage and the right-hand side gives an upper bound to this difference - twice the supremum distance between the two distribution functions F_{T_n} and Φ. Tying this back to previous sections by considering F_{T_n} the Edgeworth expansion of the cumulative distribution function we can see that the right hand side is precisely two times the expression we presented in (2.21). Thus we can say that as this distance approaches zero, so does the difference between the nominal and true coverage of the confidence interval in (2.23), which would lead to better inference of the unknown parameter θ. This concludes the theory of this thesis and we will now move on to applications.

3. Applications

In this chapter, we will look at two example distributions - the Poisson and the exponential. We will derive the cumulants for the standardized sample variables, the variance-stabilizing transformations and then the cumulants for the variance-stabilized sample variables. Having gathered the cumulants we will plug them into the Edgeworth expansion of the cumulative distribution function for each variable and, using (2.21), we will see if the VST brings the variable closer to normality in each of the sub-intervals presented in (2.22) or not.

Having gone through both of the examples and interpreted their respective results we will suggest a general improvement to the variance-stabilized statistic, apply it, and present the new results.

3.1. Poisson distribution

Let $X_{1:n} = (X_1, \dots, X_n)$ be a sample of iid $Po(\lambda)$ random variables. When referring to the underlying sample distribution we will henceforth just refer to it as X. The Poisson distribution belongs to the natural exponential family because its probability mass function can be rewritten as

$$f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} = \frac{1}{x!}\exp\{x \ln \lambda - \lambda\}.$$

We can see that this matches the criteria from Definition 2.5, hence we can directly retrieve the natural parameter as $\eta(\lambda) = \ln \lambda$ and the function $A(\eta(\lambda)) = \lambda$ which further yields $A(\eta) = e^\eta$. From this, using (2.18), we retrieve the cumulant generating function as

$$K_X(t) = A(\eta + t) - A(\eta) = e^{\ln \lambda + t} - e^{\ln \lambda} = \lambda(e^t - 1).$$

The cumulants are now rather easily derived as

$$\kappa_r(X) = K_X^{(r)}(0) = \lambda, \qquad \text{for } r = 1, 2, \dots.$$

Now defining

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\lambda}}$$

as our standardized sample variable, where $\overline{X}_n$ is the sample mean, we can obtain its third and fourth cumulants, using the result from Example 2.2, as

$$\kappa_r(Z_n) = \frac{n}{(\sqrt{n\lambda})^r}\kappa_r(X) \qquad \text{for } r = 2, 3, \dots$$

which yields

$$\begin{aligned} \kappa_1(Z_n) &= 0 \\ \kappa_2(Z_n) &= 1 \\ \kappa_3(Z_n) &= 1/\sqrt{n\lambda} \\ \kappa_4(Z_n) &= 1/(n\lambda), \end{aligned} \tag{3.1}$$

since we already know the first and second cumulants due to the standard expectation and variance of Z_n.

Remark 3.1. Together with (2.16), the cumulants from (3.1) nicely show why standardized Poisson variables converge to normality "faster" when the rate parameter is high, even though the sample size might be very small. The bigger the product $n\lambda$ gets, the smaller the correction terms in the Edgeworth expansion become - hence the CDF for the standard normal distribution will on its own approximate F_{Z_n} rather well in those cases.

Now moving on to the variance-stabilized variable, the first step is to retrieve the variance-stabilizing transformation (we already did this in Example 2.1, but for the sake of coherence we will do it again). Using (2.20) we can retrieve the VST as

$$g(\lambda) = c \int \frac{1}{\sqrt{\kappa_2(X)}} d\lambda = c\frac{\sqrt{\lambda}}{1/2} \overset{c=1/2}{=} \sqrt{\lambda},$$

and through the delta method (2.2) construct a variable

$$\begin{aligned} \sqrt{n}\left(\sqrt{\overline{X}_n} - \sqrt{\lambda}\right) &\xrightarrow{D} N(0, (g'(\lambda))^2\lambda) \\ &= N(0, 1/4), \end{aligned}$$

which yields

$$V_n = 2\left(\sqrt{n\overline{X}_n} - \sqrt{n\lambda}\right) \xrightarrow[(2.2)]{D} N(0,1)$$

as our variance-stabilized sample variable. Looking at V_n we can see that

$$n\overline{X}_n = X_1 + \ldots + X_n \sim Po(n\lambda).$$

Defining $Y = n\overline{X}_n$ we can instead look at

$$V_n = 2(\sqrt{Y} - \sqrt{n\lambda}), \tag{3.2}$$

which simplifies things a bit. If we now calculate the cumulants of $\sqrt{Y}$ we can then calculate the cumulants of V_n using the homogeneity and translational invariance of (2.10). We decide to use the moments (2.9) in order to express the cumulants. Using Definition 2.2 we get the moments of $\sqrt{Y}$ as

$$\begin{aligned} \mu_r = \mu_r(\sqrt{Y}) &= \frac{d^r}{dt^r} E\left[e^{t\sqrt{Y}}\right]\Big|_{t=0} = E\left[\frac{d^r}{dt^r} e^{t\sqrt{Y}}\right]\Big|_{t=0} \\ &= E\left[Y^{r/2}\right] = \sum_{y=0}^{\infty} y^{r/2} e^{-n\lambda} \frac{(n\lambda)^y}{y!}. \end{aligned} \tag{3.3}$$

We will have to evaluate this sum numerically for $r = 1, 2, 3, 4$ in order to get the first four cumulants, but before we do that we can construct the expressions for the cumulants of $\sqrt{Y}$ and V_n using these moments.

The cumulants of $\sqrt{Y}$ can then be expressed, using (2.9), as

$$\begin{aligned} \kappa_1(\sqrt{Y}) &= \mu_1 \\ \kappa_1(\sqrt{Y}) &= \mu_2 - \mu_1^2 \\ \kappa_1(\sqrt{Y}) &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \\ \kappa_1(\sqrt{Y}) &= \mu_4 - 4\mu_1\mu_3 + 12\mu_1^2\mu_2 - 3\mu_2^2 - 6\mu_1^4 \end{aligned}$$

from which we can retrieve the cumulants of V_n through

$$\begin{aligned} \kappa_1(V_n) = \kappa_1\left(2(\sqrt{Y} - \sqrt{n\lambda})\right) &\overset{(2.10)}{=} 2\left(\kappa_1(\sqrt{Y}) - \sqrt{n\lambda}\right) = 2\mu_1 - 2\sqrt{n\lambda} \\ \kappa_r(V_n) &\overset{r\geq 1}{=} 2^r \kappa_r(\sqrt{Y}). \end{aligned}$$

More explicitly expressed we get

$$\begin{aligned} \kappa_1(V_n) &= 2(\mu_1 - \sqrt{n\lambda}) \\ \kappa_2(V_n) &= 4(\mu_2 - \mu_1^2) \\ \kappa_3(V_n) &= 8(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3) \\ \kappa_4(V_n) &= 16(\mu_4 - 4\mu_1\mu_3 + 12\mu_1^2\mu_2 - 3\mu_2^2 - 6\mu_1^4) \end{aligned} \tag{3.4}$$

as the cumulants of the variance-stabilized variable.

We now have expressions for the cumulants of both Z_n and V_n. What is left to do is to compare the two variables using (2.21). However, before doing this, we can make a rather nice simplification. Looking at the cumulants from (3.1), the moments from (3.3) and the cumulants from (3.4) we can see that they all depend only on the product of n and λ. We could therefore introduce a new variable $\theta = n\lambda$ in order to avoid having to adjust for two parameters.

So, how do we choose which values of θ to look at? We have an infinite sum that we want to approximate numerically (see Appendix A.2.3), so we are a bit bounded by what our computer can handle. After some experimentation, we land on four different values: $\theta = 5, 10, 25, 100$. For the lowest $\theta = 5$, the error seems to be of order 10^{-3} and for the highest $\theta = 100$, the error decreases to the order 10^{-6}.

For the sub-intervals in (2.22) we get the following results:

Table 3.1.: Cumulants of Z_n and suprema differences between $F^G_{Z_n}(x)$ and $\Phi(x)$ when underlying sample distribution is Poisson

θ	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
5	0	1	0.447	0.20	0.01657	0.03111	0.03357	0.02878	0.01181
10	0	1	0.316	0.10	0.01092	0.02080	0.02240	0.01947	0.00842
25	0	1	0.200	0.04	0.00651	0.01263	0.01365	0.01206	0.00549
100	0	1	0.100	0.01	0.00310	0.00615	0.00669	0.00510	0.00285

Table 3.2.: Cumulants of V_n and suprema differences between $F^G_{V_n}(x)$ and $\Phi(x)$ when underlying sample distribution is Poisson. Green cells signify improvement over Z_n and the rest can be considered either unchanged or worse.

θ	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
5	-0.130	1.144	-0.595	1.408	0.04948	0.05359	0.01861	0.05985	0.01800
10	-0.083	1.045	-0.211	0.209	0.01820	0.02215	0.01894	0.02003	0.01128
25	-0.051	1.016	-0.110	0.049	0.00977	0.01377	0.01328	0.01272	0.00719
100	-0.025	1.004	-0.061	0.203	0.00461	0.00459	0.00833	0.01090	0.00423

For $i = 1, \dots, 4, j = 1, \dots, 9$ let i denote the rows, and j the columns. We should then compare entry (i, j) in Table 3.1 to entry (i, j) in Table 3.2. We can start by looking at the cumulants. According to the delta method (2.2) we expect to see $\kappa_1(V_n) \to 0$ and $\kappa_2(V_n) \to 1$, and this seems to be the case, which is reassuring. Looking at the third cumulants, $j = 3$, we can see that except for $\theta = 5$, the size of $\kappa_3(V_n)$ has shrunk significantly in comparison to $\kappa_3(Z_n)$ - this is a good result. However, when looking at the fourth cumulant, $j = 4$, we can see a similar increase on $\kappa_4(V_n)$ in comparison to $\kappa_4(Z_n)$

for all values on θ except $\theta = 25$ where the increase is more mild.

Moving on to the suprema we can see that there are only four cells in Table 3.2 that have smaller suprema than Table 3.1. Three of these are in the interval $I_3 = [-0.25, 0.25]$ - where the probability mass is the most dense - circa 20% of the mass lies in this interval. For $\theta = 5$ the relative difference between V_n and Z_n is more significant than in the other cases - the supremum difference of V_n has almost halved in size in comparison to Z_n. In all of the other cases, the relative difference is less significant.

In conclusion, we cannot say that the variance-stabilizing transformation yields significant improvements in the Poisson case - in the sense that the Edgeworth expansion of the cumulative distribution function from (2.14) for our variance-stabilized variable does not land closer to the standard normal CDF than before transformation. There is some evidence that points to improvements for values close to the center of the distribution, but only for very small $\theta = n\lambda$ - in which cases the exact distribution would probably be rather easy to calculate without using a normal approximation.

3.2. Exponential distribution

Let $X_{1:n}$ denote an iid sample of $Exp(\mu)$ variables with $E[X_i] = \mu$ and $\text{Var}(X_i) = \mu^2$. From here on out, let X represent the underlying sample distribution. We define our standardized variable as

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\mu}.$$

In order to retrieve the cumulants of X this time, we start from the moment-generating function (see Gut 2009, Appendix B)

$$M_X(t) = \frac{1}{1 - \mu t},$$

from which the cumulant-generating function can be retrieved as

$$K_X(t) \overset{2.3}{=} \ln(M_X(t)) = -\ln(1 - \mu t).$$

Using a standard Maclaurin expansion, this can be expanded to

$$\begin{aligned} K_X(t) &= \mu t + \mu^2 \frac{t^2}{2} + \mu^3 \frac{t^3}{3} + \mu^4 \frac{t^4}{4} + \dots \\ &= \mu t + \mu^2 \frac{t^2}{2} + 2\mu^3 \frac{t^3}{3!} + 3!\mu^4 \frac{t^4}{4!} + \dots, \end{aligned}$$

in which the r:th cumulant of X can be extracted as the coefficient preceding $t^r/r!$ as in Definition 2.4. Now using the result from Example 2.2 once again,

$$\kappa_r(Z_n) \overset{r \geq 1}{=} \frac{n}{(\mu\sqrt{n})^r} \kappa_r(X),$$

and the cumulants of Z_n evaluate to

$$\begin{aligned} \kappa_1(Z_n) &= 0 \\ \kappa_2(Z_n) &= 1 \\ \kappa_3(Z_n) &= \frac{2}{\sqrt{n}} \\ \kappa_4(Z_n) &= \frac{6}{n}. \end{aligned} \tag{3.5}$$

Notice how the cumulants of Z_n are completely independent of μ. This can be understood by rewriting the variable as

$$Z_n = \frac{\sqrt{n}}{\mu}\overline{X}_n - \sqrt{n},$$

from which using the MGF

$$M_{\frac{\sqrt{n}}{\mu}\overline{X}_n}(t) = M_{X_1+\ldots+X_n}\left(\frac{t}{\mu\sqrt{n}}\right) = \left(\frac{1}{1-\mu\frac{t}{\mu\sqrt{n}}}\right)^n = \left(\frac{1}{1-\frac{1}{\sqrt{n}}t}\right)^n,$$

we can recognize that Z_n itself is actually completely independent of μ. This is because μ is a scaling factor, and the standardization removes scale from the equation. So, for Z_n, we only really have to consider one parameter, and that is the sample size n.

Now let us move on to the variance-stabilizing transformation. Using (2.3) we can retrieve it as

$$g(\mu) = c\int \frac{1}{\sqrt{\kappa_2(X)}}d\mu = c\int \frac{1}{\mu}d\mu \overset{c=1}{=} \ln\mu,$$

which gives us the VST variable

$$V_n = \sqrt{n}(\ln\left(\overline{X}_n\right) - \ln\mu) \xrightarrow[(2.2)]{D} N(0,1).$$

Now, in order to get the cumulants of V_n we can start by looking at the distribution of $\overline{X}_n$. Through the MGF we get

$$M_{\overline{X}_n}(t) = M_{X_1+\ldots+X_n}\left(\frac{t}{n}\right) = \left(\frac{1}{1-\frac{\mu}{n}t}\right)^n,$$

which tells us that $\overline{X}_n \sim Gamma(n, \mu/n)$ where n is the shape and μ/n is a scale parameter. Let $Y = \overline{X}_n$ and denote $\theta = \mu/n$ in order to simplify notation. The MGF of $\ln Y$ can then be retrieved as

$$\begin{aligned}
M_{\ln Y}(t) &= E[e^{t\ln Y}] = E[Y^t] \\
&= \int_0^\infty y^t f_Y(y)dy \\
&= \int_0^\infty y^t \frac{1}{\Gamma(n)}y^{n-1}\frac{1}{\theta^n}e^{-y/\theta}dy \\
&= \theta^t\frac{\Gamma(n+t)}{\Gamma(n)}\int_0^\infty \frac{1}{\Gamma(n+t)}y^{n+t-1}\frac{1}{\theta^{n+t}}e^{-y/\theta}dy \\
&= \theta^t\frac{\Gamma(n+t)}{\Gamma(n)},
\end{aligned}$$

using the normalizing condition of density functions in the last step. From this we can, using Definition 2.3 get the cumulant-generating function

$$K_{\ln Y}(t) = \ln(M_{\ln Y}(t)) = t\ln\frac{\mu}{n} + \ln(\Gamma(n+t)) - \ln(\Gamma(n)),$$

from which we can tell, using Definition 2.4, that

$$\begin{aligned}
\kappa_1(\ln Y) &= \ln\frac{\mu}{n} + \frac{d}{dt}\Big[\ln(\Gamma(n+t))\Big]\Big|_{t=0} \qquad \text{and} \\
\kappa_r(\ln Y) &= \frac{d^r}{dt^r}\Big[\ln(\Gamma(n+t))\Big]\Big|_{t=0}, \qquad r>1.
\end{aligned}$$

Remark 3.2. Taking derivatives of the logarithm of the gamma function yields the polygamma function, which is defined as

$$\Psi^{(r)}(z) = \frac{d^{r+1}}{dz^{r+1}} \ln(\Gamma(z)).$$

These can be evaluated using some software of choice. In this thesis, we use the function "psi" from the R-package "pracma" in order to evaluate these derivatives.[1]

We thus get the cumulants of $\ln Y$ as

$$\begin{aligned}
\kappa_1(\ln Y) &= \ln \frac{\mu}{n} + \Psi^{(0)}(n) \\
\kappa_2(\ln Y) &= \Psi^{(1)}(n) \\
\kappa_3(\ln Y) &= \Psi^{(2)}(n) \\
\kappa_4(\ln Y) &= \Psi^{(3)}(n),
\end{aligned}$$

from which we can, using (2.10), express the cumulants of V_n as

$$\begin{aligned}
\kappa_1(V_n) &= \sqrt{n}(\kappa_1(\ln Y) - \ln \mu) = \sqrt{n}(\Psi^{(0)}(n) - \ln n) \\
\kappa_r(V_n) &= n^{r/2}\kappa_r(\ln Y), \qquad r > 1
\end{aligned}$$

which yields, more explicitly stated

$$\begin{aligned}
\kappa_1(V_n) &= \sqrt{n}(\Psi^{(0)}(n) - \ln n) \\
\kappa_2(V_n) &= n\Psi^{(1)}(n) \\
\kappa_3(V_n) &= n^{3/2}\Psi^{(2)}(n) \\
\kappa_4(V_n) &= n^2\Psi^{(3)}(n).
\end{aligned} \tag{3.6}$$

As we can see, both the cumulants of Z_n (3.5) and V_n (3.6) are completely independent of μ and only depend on the sample size n. This allows us to focus solely on the sample size parameter n when performing evaluations.

Now, we turn to the comparison of the Edgeworth expansions (2.14) of the distribution functions of Z_n and V_n. We first have to decide on what values of n to look at. This time, since we have the cumulants of the variance-stabilized variable as closed analytical expressions - as some constant multiplied with a polygamma-function evaluated at n - we are not held back by computational restrictions in the same way as in the Poisson case. We are therefore allowed to try out some pretty hefty numbers on n, which might be interesting. After some light experimentation we decide on the values $n = 10, 10^2, 10^3, 10^4$. For our sub-intervals we use the same as in the Poisson case, the ones presented in (2.22).

[1] https://search.r-project.org/CRAN/refmans/pracma/html/psi.html

Table 3.3.: Cumulants of Z_n and suprema differences between $F^G_{Z_n}(x)$ and $\Phi(x)$ when underlying sample distribution is exponential

n	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
10	0	1	0.632	0.6000	0.02650	0.04830	0.05802	0.05302	0.01982
10^2	0	1	0.200	0.0600	0.00656	0.01251	0.01380	0.01244	0.00553
10^3	0	1	0.063	0.0060	0.00193	0.00384	0.00422	0.00383	0.00183
10^4	0	1	0.020	0.0006	0.00060	0.00121	0.00133	0.00121	0.00059

Table 3.4.: Cumulants of V_n and suprema differences between $F^G_{V_n}(x)$ and $\Phi(x)$ when underlying sample distribution is exponential. Green cells signify improvement over Z_n and the rest can be considered either unchanged or worse.

n	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
10	-0.161	1.05166	-0.349	0.2320	0.03116	0.04293	0.04049	0.03979	0.02237
10^2	-0.050	1.00502	-0.101	0.0203	0.00859	0.01335	0.01332	0.01316	0.00771
10^3	-0.016	1.00050	-0.032	0.0020	0.00261	0.00421	0.00421	0.00419	0.00252
10^4	-0.005	1.00005	-0.010	0.0002	0.00082	0.00133	0.00133	0.00133	0.00081

Comparing Table 3.3 and Table 3.4, we can start by looking at the cumulants. From Table 3.4 we see that $\kappa_1(V_n)$ and $\kappa_2(V_n)$ appear to tend to 0 and 1 respectively as n gets larger - which somewhat confirms that we have used the delta method correctly. Moving on to the third and fourth cumulants we can see that they have significantly decreased in absolute value post-VST. For the third cumulant, for all n, the cumulants of V_n are close to half the size of the third cumulant of Z_n. The fourth cumulant has seen an even more significant decrease - the fourth cumulants of V_n are close to one third of the size of the fourth cumulant of Z_n, for all n. This we believe bodes well for the variance-stabilized variable, but when actually comparing suprema between the two tables we can see that the results are pretty underwhelming.

For $x \in I_1, I_5$ - far out in the tails - Z_n actually lands closer to the normal distribution than V_n for all n. This, however, only accounts for roughly 10% of the probability mass of a normal distribution. Looking at the 90% that remain we can see that for $n = 10$ V_n actually lands closer to normality than Z_n. Right in the middle, $x \in I_3$, we can see that V_n is actually closer to normality for $n = 10, 10^2, 10^3$ and identical to Z_n for $n = 10^4$. However, only $n = 10$ see a significant improvement whereas for $n = 10^2, 10^3$ the improvement is very mild.

These results are somewhat similar to the Poisson distribution, albeit a bit more tilted towards improvement for the variance-stabilized variable. We only really see improvement right in the middle of the distribution, at $I_3 = [-0.25, 0.25]$, and for a very low $n = 10$. We do however see significant improvements on the third and fourth cumulants after variance-stabilization. So why do these improvements on the cumulants not permeate through the Edgeworth expansion? It is hard to tell from just looking at these numbers but one such reason could be that the standardized mean and variance of Z_n just weighs heavier in favor of normality than improvements on the third and fourth cumulants. Recall from Remark 2.6 how three entire coefficients out of the eight in total evaluate to zero when the first cumulant is 0 and the second is 1. This is already a pretty powerful attribute, that Z_n possess. All in all, we would probably not recommend using the variance-stabilizing transformation in the exponential case either. Not in its current form at least.

3.3. Improving the VST

Reflecting on the results of the Poisson distribution case, it might not be that unreasonable that the VST did not improve convergence towards normality since we only saw an improvement on the third cumulant, whereas the fourth cumulant was significantly worsened - and standardization of mean and variance was lost - by the transformation. In the exponential case, however, both the third and fourth cumulants saw significant improvements, but still: underwhelming results. This has lead us to believe that true standardization weighs far heavier as a component than improvements on skewness and kurtosis when trying to approach normality.

So let us just standardize the variance-stabilized statistic. We have

$$E[V_n] = \kappa_1(V_n),$$
$$\mathrm{Var}(V_n) = \kappa_2(V_n).$$

Hence, we can define a new variable

$$V_n^{std} = \frac{V_n - \kappa_1(V_n)}{\sqrt{\kappa_2(V_n)}}.$$

The cumulants would now theoretically, using (2.10), evaluate to

$$\begin{aligned} \kappa_1(V_n^{std}) &= 0 \\ \kappa_2(V_n^{std}) &= 1 \\ \kappa_3(V_n^{std}) &= \frac{\kappa_3(V_n)}{(\kappa_2(V_n))^{3/2}} \\ \kappa_4(V_n^{std}) &= \frac{\kappa_3(V_n)}{(\kappa_2(V_n))^2}. \end{aligned} \tag{3.7}$$

We say "theoretically" here because in the Poisson case we have actually not been able to derive the exact cumulants for the variance-stabilized variable, only approximations. We should therefore take care to use these approximated values when creating this new, standardized, variance-stabilized variable V_n^{std}.

Looking back at Table 3.2 (VST Poisson) and Table 3.4 (VST exponential), we can see that the second cumulant, $\kappa_2(V_n)$, is actually approaching 1 from the positive direction in both cases. Now looking at the third and fourth cumulants of (3.7) this leads us to believe that, apart from regaining the seemingly powerful property of standardized expectation and variance, we should see even further improvements on skewness and kurtosis for V_n^{std}. As long as this second cumulant of the variance-stabilized variable approaches 1 from the positive direction, this should hold for any variance-stabilized variable, with no regards to the underlying sample distribution.

3.3.1. Standardized VST Poisson distribution

Now, using (3.7), we compile the results for V_n^{std} in the Poisson distribution case. We show the results for Z_n once again in order to make comparing them easier.

Now comparing Z_n to V_n^{std}, the results are as follows ($\theta = n\lambda$):

Cumulants of Z_n and suprema differences between $F_{Z_n}^{G}(x)$ and $\Phi(x)$ when underlying sample distribution is Poisson

θ	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
5	0	1	0.447	0.20	0.01657	0.03111	0.03357	0.02878	0.01181
10	0	1	0.316	0.10	0.01092	0.02080	0.02240	0.01947	0.00842
25	0	1	0.200	0.04	0.00651	0.01263	0.01365	0.01206	0.00549
100	0	1	0.100	0.01	0.00310	0.00615	0.00669	0.00600	0.00285

Table 3.5.: Cumulants of V_n^{std} and suprema differences between $F_{V_n^{std}}^{G}(x)$ and $\Phi(x)$ when underlying sample distribution is Poisson. Green cells signify improvement over Z_n and the rest can be considered either unchanged or worse.

θ	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
5	0	1	-0.486	1.075	0.02335	0.06429	0.06427	0.03643	0.01790
10	0	1	-0.197	0.191	0.00592	0.01495	0.01523	0.01132	0.00683
25	0	1	-0.107	0.047	0.00310	0.00699	0.00737	0.00627	0.00341
100	0	1	-0.061	0.202	0.00261	0.00729	0.00671	0.00449	0.00213

As can be seen from Table 3.5, there is a lot more green now, in comparison to Table 3.2. Even though the fourth cumulant is still a lot worse for V_n^{std} than Z_n whereas the improvements on the third cumulant is about the same as it was for the non-standardized, variance-stabilized V_n, the improvement seems significant. This tells us that standardized expectation and variance really does mean a lot for convergence towards normality. Furthermore, these results somewhat indicate that improvement on the skewness weigh heavier than improvements on kurtosis. Maybe the lower the order of the cumulant, the more it affects the properties of the distribution? If such is the case, it would be nice if one could formulate it more concretely. We will return to this in the final chapter.

Either way, in the Poisson case, standardizing the variance-stabilized variable has definitely shown some evidence of improvement. If one is working with a $\theta = n\lambda$ of either 10 or 25 we would at least not feel too bad mentioning the standardized variance-stabilized transformation as an option. We are, however, a bit weary of the results on the fourth cumulant for both versions of the variance-stabilized statistic. Maybe one should look to other types of transformations if the goal is to improve inference for Poisson variables. This, we will also return to in the final chapter.

3.3.2. Standardized VST exponential distribution

Once again, let us compile the results for V_n^{std} using (3.7) and compare the results to Z_n when the underlying sampling distribution is exponential.

Cumulants of Z_n and suprema differences between $F^G_{Z_n}(x)$ and $\Phi(x)$ when underlying sample distribution is exponential

n	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
10	0	1	0.632	0.6000	0.02650	0.04830	0.05802	0.05302	0.01982
10^2	0	1	0.200	0.0600	0.00656	0.01251	0.01380	0.01244	0.00553
10^3	0	1	0.063	0.0060	0.00193	0.00384	0.00422	0.00383	0.00183
10^4	0	1	0.020	0.0006	0.00060	0.00121	0.00133	0.00121	0.00059

Table 3.6.: Cumulants of V_n^{std} and suprema differences between $F^G_{V_n^{std}}(x)$ and $\Phi(x)$ when underlying sample distribution is exponential. Green cells signify improvement over Z_n and the rest can be considered either unchanged or worse.

n	κ_1	κ_2	κ_3	κ_4	$\sup_{x\in I_1}$	$\sup_{x\in I_2}$	$\sup_{x\in I_3}$	$\sup_{x\in I_4}$	$\sup_{x\in I_5}$
10	0	1	-0.324	0.2098	0.00906	0.02254	0.02443	0.02087	0.01162
10^2	0	1	-0.100	0.0201	0.00287	0.00615	0.00675	0.00605	0.00312
10^3	0	1	-0.032	0.0020	0.00093	0.00191	0.00211	0.00190	0.00095
10^4	0	1	-0.010	0.0002	0.00030	0.00060	0.00066	0.00060	0.00030

Green all over! Surely, Table 3.6 provides some pretty strong evidence in favor of the standardized variance-stabilized sample variable V_n^{std} approaching normality faster than the simply standardized variable Z_n, when the underlying sample distribution is exponential. This also furthers our current hypothesis that standardization of mean and variance weigh heavier than improvements on the third and fourth cumulants. Having the first two cumulants now standardized this allows for the improvements on skewness and kurtosis to really shine through. Using this transformation in order to improve inference when the underlying sample distribution is exponential might be a rather viable option.

4. Conclusions and Discussion

4.1. About the theoretical framework

The theory we have presented should in a sense serve as a motivation for why the cumulants are of interest when studying convergence towards normality from a purely theoretical perspective (and perhaps convergence in a more general sense as well). The cumulants depend on the cumulant-generating function and the cumulant-generating function in turn depends on the moment-generating (or characteristic) function. So in a sense, it all comes down to how the MGF (or characterstic function) is affected by the given transformation. We know that a distribution is uniquely determined by its MGF $M : \mathbb{R} \to (0, \infty)$ (if it exists) and the CGF is $\log M : (0, \infty) \to \mathbb{R}$. This yields a one-to-one correspondence between the generating functions and no information is lost. Depending on context, one could choose to study whichever generating function that yields the easiest calculations.

4.1.1. Benefits of cumulants

There are, however, some noteworthy benefits of studying the cumulant-generating function that one should be aware of when choosing which generating function to work with.

Recall from introductory courses in probability that

$$E\left[\sum_i X_i\right] = \sum_i E[X_i]$$

always holds true, and that in the iid case

$$\mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}(X_i).$$

The cumulants are in a sense a generalization of this property. The first cumulant is the expectation, the second is the variance and the r:th cumulant of a sum of iid random variables is

$$\kappa_r\left(\sum_i X_i\right) \overset{(2.10)}{=} \sum_i \kappa_r(X_i).$$

They are cumulative, hence their name. Remember also how

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X).$$

This property is also generalized, for $r = 2, 3, \ldots$ as

$$\kappa_r(aX + b) = a^r \kappa_r(X).$$

Assuming one were to choose cumulants as means of studying convergence towards normality: the cumulants of the standard normal distribution are $\kappa_1 = 0$, $\kappa_2 = 1$ and $\kappa_3 = \kappa_4 = \kappa_5 = \ldots = 0$, and no other distribution holds this property. Therefore, if one were to show that the rate at which the cumulants converge to the set of the standard normal cumulants increases after some transformation, for all cumulants, it should follow that the rate of convergence towards normality increases as well. If this is possible, we would in a sense be isolating the part of the problem that actually matters. Using the MGF this would involve the problem of solving integrals (in the continuous case) in the form of

$$M_{g(X)}(t) = E[e^{g(X)t}] = \int_{-\infty}^{\infty} e^{g(x)t} f_X(x)dx,$$

for some random variable X and transformation g. Depending on g this might be really hard, or impossible, to solve analytically and (like in the Poisson case) numerical methods would have to be applied in order to approximate the integral. There might be some tricks one could try using Jensen's inequality (see Held and Bové 2020, p. 354) but this would depend on g. Assume

$$h(x) = e^{g(x)t}$$

is a concave function, $h''(x) \leq 0$. It then follows from Jensen's inequality that

$$E[h(X)] = M_{g(X)}(t) \leq e^{g(E[X])t}.$$

Now taking the logarithm of both sides yields

$$K_{g(X)}(t) \leq g(E[X])t,$$

an upper bound for the cumulant-generating function of the transformed variable that seems fairly easy to work with. Of course, in order to find the cumulants one would have to take derivatives of this, which probably breaks the inequality. Also, the cumulant-generating function is defined on the whole real line - hence, an upper bound is not enough. We mention Jensen's inequality because through it we are able to harness the inverse relationship between the logarithm and the exponential, thus simplifying the expressions. For further work, it would be interesting to see if this would and could be useful from an analytical perspective.

4.1.2. The Edgeworth expansion as a practical tool

Using the Edgeworth expansion as a means of studying convergence in a more practical sense requires making a number of decisions. How many cumulants should we choose? The more we choose, the more complex of an expression we have to work with. The coefficients increase in number, but also the terms within each coefficient increase, and the Hermite polynomials increase in order as well. Within this, there lies also the problem of deciding how many terms of the Edgeworth expansion to include, which depends on how many terms from the Maclaurin expansion of the moment-generating function we choose to include. Making informed decisions throughout all of this requires a pretty sophisticated understanding of mathematical analysis and probability theory, making it less accessible as a practical tool. If one manages to convince themselves of the validity of the Edgeworth expansion, it might be fruitful to try and find more direct methods of studying convergence from a theoretical perspective, possibly using only cumulants. This might, however, turn out to be challenging.

Briefly returning to the speculation we mentioned in the second to last paragraph of Section 3.3.1 - that our results seem to indicate that the significance of the cumulants seem to decrease as their orders increase, in terms of how much of an effect they have on convergence towards normality. Actually, we have already given an answer to this, failing to mention the consequences. Remembering the result from Example 2.2, that

$$\kappa_r(Z_n) = \frac{n}{(\sigma\sqrt{n})^r}\kappa_r(X), \qquad r > 1$$

it is evident that $\kappa_r(Z_n) = \mathcal{O}(n^{1-r/2})$ as long as $\kappa_r(X)$ is finite (which is not necessarily the case). This tells us that we should expect higher order cumulants to have less of an impact on convergence towards normality simply as a consequence of them being smaller and therefore having less of an impact in the Edgeworth expansion. In our generalization of the Edgeworth expansion (2.14), we loosened the assumption on the cumulants to $\kappa_r(T_n) - \mathbb{I}_{r=2} = \mathcal{O}(n^{-1/2})$ for $r = 1, 2, 3, 4$, but we do in fact hope for a similar behaviour from the cumulants of the not necessarily standardized T_n as well, due to it being asymptotically standard normal. Now looking back to the coefficients $A_1, \ldots, A_8$ from (2.11), we can see that built into the Edgeworth expansion is not only a weighing of the cumulants but also the interactions between them. It is probably possible to intuit effects on convergence simply by studying how the cumulants change through some transformation, but actually giving a number to those changes seems challenging to do without using the Edgeworth expansion.

4.1.3. Neglected downsides of the Edgeworth expansion

A fact that we have neglected to mention about Edgeworth expansions, but that may have affected the outcome, is that the rate of convergence in the tails of the distribution is usually slower than it is in the centre, when using Edgeworth expansions. In order to adjust for this fact one may use something called saddle-point approximations (see DasGupta 2008, chapter 14). We have, however, not had the time to delve into this during the work on this thesis. It should however be noted, that the results we get far out in the tails in our Edgeworth expansions may be inaccurate.

4.1.4. The Berry-Esseen bound

When applying the central-limit theorem for some variable $T_n \xrightarrow{D} N(0,1)$ we are using the approximation

$$F_{T_n}(x) = \mathbb{P}(T_n \leq x) \approx \mathbb{P}(Z \leq x) = \Phi(x)$$

where $Z \sim N(0,1)$. Hence, the error of our approximation would be

$$\left| F_{T_n}(x) - \Phi(x) \right|,$$

which is very similar to what we have seen earlier. The Berry-Esseen bound states that there exists a universal constant C such that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq x \right) - \Phi(x) \right| \leq \frac{C\beta_3}{\sigma^3 \sqrt{n}},$$

where $\overline{X}_n$ is the mean of an iid sample $\overline{X}_{1:n}$, $\mu = E[X_i]$, $\sigma^2 = \text{Var}(X_i)$ and $\beta_3 = E[(X_i - \mu)^3] < \infty$ (see DasGupta 2008, p. 142). Looking a bit closer at this we can see that $\beta_3/(\sigma^3\sqrt{n})$ is actually the same as the $\rho_3/\sqrt{n}$ coefficient present in the first error term in our initial presentation of the Edgeworth expansion (2.5). Hence, it is possible to give an upper bound for the error of the approximation using only the third cumulant and this universal constant C. One drawback of this bound however, as opposed to the generalization of the Edgeworth expansion we have presented is that it requires our sample variable to be standardized, which is a property we usually lose when using the simple version of the variance-stabilizing transformation. It would, however, be interesting to check if this is applicable to the standardized VST.

There exist modifications of the Berry-Esseen bound that loosen the restriction on standardization, but in such cases the constant C is no longer universal and harder to find (see

Theorem 11.6 DasGupta 2008, p. 146). These bounds are, however, valuable in terms of understanding the rate of convergence. Loosely speaking, the Berry-Esseen bound can be considered a precursor to the Edgeworth expansion.

4.2. About variance-stabilizing transformations

As we saw in the previous chapter, standardization of the variance-stabilized variable yielded some pretty good results. This method, however, requires us to first calculate the cumulants of the variance-stabilized statistic and then use these to re-standardize the variable and calculate the new cumulants. Seemingly, this does not require a lot of extra work - see (3.7). The homogeneity (2.10) of cumulants allows us to derive these new cumulants easily. It would, however, be very nice if one could pre-adjust the VST such that it accounts for these errors right from the start.

4.2.1. Errors introduced by the VST

What we are actually doing when applying the variance-stabilizing transformation is that we are making the first-order term in the asymptotic expansion of the variance of the statistic independent of the unknown parameter (see "b" in Theorem 3.9 DasGupta 2008, p. 44) - we are making this term a constant. This means that higher order terms could still very much be dependent on the unknown parameter and such, the variance is not really a constant for finite samples. Furthermore we are actually introducing some bias when applying the VST, usually of the second order. One option would therefore be to try and adjust for these second-order errors preemptively.

From the literature (see DasGupta 2008, p. 54-55), we can see that perturbing the transformation by some constant can achieve either second-order bias correction or second-order variance-stabilization - but apparently no set of constants can achieve both simultaneously. Because of this, we deem the method of perturbing the transformation with a constant as not fruitful enough. The re-standardization method would then be a better option. We have not been able to find clear evidence that simultaneous second-order bias-correction and variance-stabilization is possible at all. Apparently, second-order bias-correction on its own is rather fruitful and we will now present a general method for achieving it.

4.2.2. Second-order bias correction of the VST

Consider, for some statistic T_n, the second order Taylor expansion around $E[T_n] = \mu$, where μ is the expected value of the statistic,

$$\begin{aligned} E[g(T_n)] &\approx E\left[g(\mu) + g'(\mu)(T_n - \mu) + \frac{1}{2}g''(\mu)(T_n - \mu)^2\right] \\ &= g(\mu) + \frac{1}{2}g''(\mu)\mathrm{Var}(T_n). \end{aligned}$$

We want a method of adjusting the VST such that this second term cancels.

Example 4.1. In the Poisson distribution case we have

$$\begin{aligned} E[\overline{X}_n] &= \lambda \\ \mathrm{Var}(\overline{X}_n) &= \frac{\lambda}{n} \\ g(\lambda) &= \sqrt{\lambda}, \end{aligned}$$

which gives us

$$E\left[\sqrt{\overline{X}_n}\right] = \sqrt{\lambda} - \frac{1}{8n\sqrt{\lambda}} + \mathcal{O}(n^{-2}),$$

from which we can tell that for small n, λ this second-order bias term is certainly not negligible.

Suppose $T_n = T_n(X_{1:n})$ is an estimate of an unknown parameter θ, where $X_{1:n}$ is an iid sample of random variables. If we can then find an expansion

$$E[g(T_n)] = g(\theta) + \frac{1}{n}\left(g'(\theta)b(\theta) + \frac{1}{2}g''(\theta)\sigma^2(\theta)\right) + \mathcal{O}(n^{-2}),$$

where $\sigma^2(\theta)$ is the variance of T_n and g is our variance-stabilizing transformation - then, under "various conditions", we can construct a second-order bias-corrected VST as

$$h(T_n) = g(T_n) + \frac{\sigma'(T_n)}{2} - \frac{b(T_n)}{\sigma(T_n)}$$

(see DasGupta 2008, p. 55). Apparently, this second-order bias-correction of the VST usually leads to better inference than using the regular VST (see Remark DasGupta 2008, p. 57). There appear to exist similar methods for second-order variance-stabilization but we have not been able to find them following the references in the literature (see DasGupta 2008, p. 55 for such a reference), and like we stated in the last section - we do not know if second-order bias-correction and variance-stabilization is achievable simultaneously.

4.2.3. Some comments about Poisson and exponential VST

Were we to speculate as to why the VST performed much better in the exponential case than in the Poisson - we would point to the fact that the *log*-transformation transforms the support of the statistic from the positive real line to the entire real line, whereas the square-root transformation does not. This means that the *log*-transformed statistic already shares the same support as the normal distribution, whereas the square-root does not.

Furthermore, the square-root transformation "squishes" the relative distance to the mean. Say we have a $Po(100)$ variable and we have one observation from it: 144. The standardized observation would now be $44/10 = 4.4$ - a pretty extreme observation. Now applying the square-root (and multiplying by 2 as in (3.2)), we get (assuming the mean is roughly $\sqrt{100}$ after transformation) $2(12 - 10) = 4$ - not as extreme as before transformation. In this sense, by applying the square-root transformation we are making extreme observations less extreme - we are increasing kurtosis, which is precisely what we see from Table 3.2. Hence, the VST interacts somewhat poorly with the actual distribution of the data in the Poisson case. We should take this as a lesson that one should probably not (as we have done in this thesis) mindlessly rely on the notion that a VST is always suitable. We should instead examine it critically before applying it and form an idea of what to expect. If the transformation seems to increase any of the lower order (1 to 4) cumulants, there may exist other, more appropriate transformations that one could use instead. For example, it would be interesting to see how the *log*-transform would perform on the Poisson data - with re-standardization applied.

4.2.4. Concluding words on the VST

Although we haven't really tried out the variance-stabilizing transformation that thoroughly (only two distributions) in this thesis, the literature and our results tell us that applying the VST in its most simple form, as in (2.3), usually does not perform very well, due to the introduction of a bias and the variance (ironically enough) not really being a constant for finite samples - "a bias correction is the least one should do" (DasGupta 2008, p. 59).

It should be noted that for the area close to the centre of the distribution we have seen some evidence in favor of the simple VST. However, these improvements are only noteworthy for very small sample sizes - $n = 5, 10$ in the Poisson case and $n = 10$ in the exponential. Surely, in both of these cases, using the exact distribution in order to infer about the unknown parameter is entirely possible - hence there might be no need to apply the central limit theorem.

Regarding the third and fourth cumulants (skewness and kurtosis correction factors) we have seen mixed results. Disregarding the very small sample size of $n = 5$ in the Poisson case, the skewness saw pretty significant improvements in both the Poisson and exponential distribution cases. On the fourth cumulant, however, we are back to mixed results. In the exponential case the improvement after variance-stabilization was significant for all tested sample sizes, while in the Poisson case the opposite was true. At the very least, we can say that the VST effect on the kurtosis correction term seems to be very much dependent on the properties of the actual transformation and how it interacts with the distribution data. The same might be true for the skewness. One should not expect the variance-stabilizing transformation to always produce good results on skewness and kurtosis. Simply because a transformation stabilizes the variance for some distribution, it does not mean that it has a beneficial impact on the rest of the cumulants.

Standardizing the mean and variance for the VST, however, seems to produce some positive results - very positive in the exponential distribution case. From this we have gathered that lower order cumulants are of greater importance than higher and one should be aware of this when applying transformations. If re-standardization is possible one could then focus on finding transformations that lower skewness and kurtosis. All in all, we believe that the standardized VST is a viable choice of transformation if the underlying sampling distribution is exponential, and in the Poisson case, we believe that one should probably look to other transformations.

4.3. Further work

For further work in a more general sense we suggest going down the route of simulation or using real data as a means of studying how transformations affect convergence. Having seen how great of an effect standard expectation and variance seem to have on convergence we would advise to always re-standardize after transforming the data. When starting from data, this should be rather straightforward in the sense that you would just apply the transformation, estimate mean and variance (or calculate exactly if possible), and re-standardize.

For further work within the framework that we have presented we recommend studying a wider range of transformations. Staying on the VST it would be interesting to study more heavy-tailed distributions, possibly ones with infinite moments and see what the effects of variance-stabilization would be in those cases. Furthermore, we could leave the VST entirely and study a wide range of transformations and distributions in general. As long as the inferential problem is univariate and the asymptotic distribution of the statistic is normal, the framework in this thesis would allow us to study its convergence.

A. Appendix

A.1. Coefficients of the generalized Edgeworth expansion

The three-term Maclaurin expansion

$$e^x = 1 + x + \frac{x^2}{2},$$

where

$$x = \kappa_1 t + (\kappa_2 - 1)\frac{t^2}{2} + \kappa_3 \frac{t^3}{6} + \kappa_4 \frac{t^4}{24}$$

is

$$1 + \left(\kappa_1 t + (\kappa_2 - 1)\frac{t^2}{2} + \kappa_3 \frac{t^3}{6} + \kappa_4 \frac{t^4}{24}\right) + \frac{1}{2}\left(\kappa_1 t + (\kappa_2 - 1)\frac{t^2}{2} + \kappa_3 \frac{t^3}{6} + \kappa_4 \frac{t^4}{24}\right)^2.$$

First expanding the squared term and gathering as coefficients of t we get

$$\frac{x^2}{2} = \frac{\kappa_1^2}{2} t^2 + \frac{\kappa_1(\kappa_2 - 1)}{2} t^3 + \left(\frac{(\kappa_2 - 1)^2}{8} + \frac{\kappa_1 \kappa_3}{6}\right) t^4 + \left(\frac{\kappa_1 \kappa_4}{24} + \frac{(\kappa_2 - 1)\kappa_3}{12}\right) t^5 + \left(\frac{\kappa_3^2}{72} + \frac{(\kappa_2 - 1)\kappa_4}{48}\right) t^6 + \frac{\kappa_3 \kappa_4}{144} t^7 + \frac{\kappa_4^2}{1152} t^8.$$

Now adding $1 + x$ we get

$$\begin{aligned} 1 + x + \frac{x^2}{2} = 1 + \kappa_1 t + \frac{1}{2}(\kappa_1^2 + \kappa_2 - 1)t^2 + \left(\frac{\kappa_1(\kappa_2 - 1)}{2} + \frac{\kappa_3}{6}\right) t^3 \\ + \left(\frac{(\kappa_2 - 1)^2}{8} + \frac{\kappa_1 \kappa_3}{6} + \frac{\kappa_4}{24}\right) t^4 + \left(\frac{\kappa_1 \kappa_4}{24} + \frac{(\kappa_2 - 1)\kappa_3}{12}\right) t^5 \\ + \left(\frac{\kappa_3^2}{72} + \frac{(\kappa_2 - 1)\kappa_4}{48}\right) t^6 + \frac{\kappa_3 \kappa_4}{144} t^7 + \frac{\kappa_4^2}{1152} t^8. \end{aligned}$$

Denoting the coefficient in front of t^r by A_r this now yields

$$\begin{aligned} &A_1 = \kappa_1, \quad A_2 = \frac{1}{2}(\kappa_1^2 + \kappa_2 - 1), \quad A_3 = \frac{\kappa_3}{6} + \frac{(\kappa_2 - 1)\kappa_1}{2}, \\ &A_4 = \frac{(\kappa_2 - 1)^2}{8} + \frac{\kappa_1 \kappa_3}{6} + \frac{\kappa_4}{24}, \quad A_5 = \frac{\kappa_1 \kappa_4}{24} + \frac{(\kappa_2 - 1)\kappa_3}{12}, \\ &A_6 = \frac{\kappa_3^2}{72} + \frac{(\kappa_2 - 1)\kappa_4}{48}, \quad A_7 = \frac{\kappa_3 \kappa_4}{144}, \quad A_8 = \frac{\kappa_4^2}{1152}, \end{aligned}$$

which we can see match the coefficients of (2.11).

A.2. R code

A.2.1. The generalized Edgeworth expansion

Below is the function used to evaluate the generalized version, (2.14), of the Edgeworth expansion of the cumulative distribution function

```
# Tn_cums a vector of the first four cumulants of Tn
# x is value to be evaluated
ee_gen <- function(Tn_cumulants, x) {
  cdf <- pnorm(x) - dnorm(x)*(
    # A_1 * H_0
    (Tn_cums[[1]]) * 1 +
    # A_2 * H_1
    ((Tn_cums[[1]]**2 + Tn_cums[[2]] - 1) / 2) * x +
    # A_3 * H_2
    ((Tn_cums[[3]]/6) + ((Tn_cums[[2]] - 1)*Tn_cums[[1]])/2) *
        (x**2 - 1) +
    # A_4 * H_3
    (((Tn_cums[[2]]-1)**2)/8 + (Tn_cums[[1]]*Tn_cums[[3]])/6 +
        Tn_cums[[4]]/24)*(x**3 - 3*x) +
    # A_5 * H_4
    ((Tn_cums[[1]]*Tn_cums[[4]])/24 + ((Tn_cums[[2]] - 1) *
        Tn_cums[[3]])/12) * (x**4 - 6*x**2 + 3) +
    # A_6 * H_5
    ((Tn_cums[[3]]**2)/72 + ((Tn_cums[[2]] - 1)*Tn_cums[[4]])/48) *
        (x**5 - 10*x**3 + 15*x) +
    # A_7 * H_6
    ((Tn_cums[[3]]*Tn_cums[[4]])/144) * (x**6-15*x**4+45*x**2-15) +
    # A_8 * H_7
    ((Tn_cums[[4]]**2)/1152) * (x**7 - 21*x**5 + 105*x**3 - 105*x)
  )
  return(cdf)
}
```

A.2.2. Example usage of the generalized Edgeworth expansion

Below is some example code showcasing how the generalized Edgeworth expansion was used in our evaluations on the interval $I = [-5, 5]$.

```
M <- 10000
k <- 5
x <- seq(-k, k, length.out = M)
# CDF of Zn
F_Zn <- map_dbl(x, ~ee_gen(Zn_cumulants, .))
```

A.2.3. The moments of $\sqrt{Po(n\lambda)}$

Below is the function used in order to evaluate the moments of $\sqrt{Y}$ in (3.3). Here, we realized that for the highest value on $\theta = n\lambda = 100$, including the 150 first terms of the sum left an error approximately of size 10^{-6}. It was not possible to go much higher than this since the factorial factor became too large for our computer to handle.

```
# returns the r:th moment of sqrt(Po(theta=n*lambda))
sqrtPoisson_moments <- function(r, theta) {
  s = 0
  # series approximation of raw moments of sqrt(Po(theta))
  for (i in seq(0:149)) {
    term = i**(r/2) * (theta)**i / factorial(i)
    s = s + term
  }
  s = s * exp(-theta)
  return(s)
}
```

Bibliography

Barndorff-Nielsen, Ole E and David Roxbee Cox (1989). *Asymptotic techniques for use in statistics*. Vol. 11. Springer.

DasGupta, Anirban (2008). *Asymptotic theory of statistics and probability*. Vol. 180. Springer.

Gut, Allan (2009). *An Intermediate Course in Probability*. 2nd. Springer Publishing Company, Incorporated. ISBN: 1441901612.

Held, Leonhard and Daniel Sabanés Bové (2020). "Likelihood and Bayesian Inference". In: *Statistics for Biology and Health. Springer, Berlin, Heidelberg,*