

On the Gaussian Approximation to Bayesian Posterior Distributions

Christoph Fuhrmann¹, Hanns-Ludwig Harney², Klaus Harney³, Andreas Müller^{4,*}

¹School of Education, Institute for Educational Research, University of Wuppertal, D-42097 Wuppertal, Germany

²Max-Planck-Institute for Nuclear Physics, Postfach 103980, D-69029 Heidelberg, Germany

³Faculty of Philosophy and Educational Science, Institute for Educational Research, Ruhr-Universität, D-44801 Bochum, Germany

⁴Faculty of Sciences/Physics Department and Institut Universitaire de Formation des Enseignants, University of Geneva, CH - 1211 Genève 4, Switzerland

Received January 27, 2021; Revised April 21, 2021; Accepted May 16, 2021

Cite This Paper in the following Citation Styles

(a): [1] Christoph Fuhrmann, Hanns-Ludwig Harney, Klaus Harney, Andreas Müller, "On the Gaussian Approximation to Bayesian Posterior Distributions," *Mathematics and Statistics*, Vol.9, No.4, pp. 535-551, 2021. DOI: 10.13189/ms.2021.090413

(b): Christoph Fuhrmann, Hanns-Ludwig Harney, Klaus Harney, Andreas Müller, (2021). *On the Gaussian Approximation to Bayesian Posterior Distributions*. *Mathematics and Statistics*, 9(4), 535-551. DOI: 10.13189/ms.2021.090413

Copyright ©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The present article derives the minimal number N of observations needed to approximate a Bayesian posterior distribution by a Gaussian. The derivation is based on an invariance requirement for the likelihood $p(x|\xi)$. This requirement is defined by a Lie group that leaves the $p(x|\xi)$ unchanged, when applied both to the observation(s) x and to the parameter ξ to be estimated. It leads, in turn, to a class of specific priors.

In general, the criterion for the Gaussian approximation is found to depend on (i) the Fisher information related to the likelihood $p(x|\xi)$, and (ii) on the lowest non-vanishing order in the Taylor expansion of the Kullback-Leibler distance between $p(x|\xi)$ and $p(x|\xi^{\text{ML}})$, where ξ^{ML} is the maximum-likelihood estimator of ξ , given by the observations x .

Two examples are presented, widespread in various statistical analyses. In the first one, a chi-squared distribution, both the observations x and the parameter ξ are defined all over the real axis. In the other one, the binomial distribution, the observation is a binary number, while the parameter is defined on a finite interval of the real axis. Analytic expressions for the required minimal N are given in both cases. The necessary N is an order of magnitude larger for the chi-squared model (continuous x) than for the binomial model (binary x). The difference is traced back to symmetry properties of the likelihood function $p(x|\xi)$.

We see considerable practical interest in our results since the normal distribution is the basis of parametric methods of applied statistics widely used in diverse areas of research (education, medicine, physics, astronomy etc.). To have an analytical criterion whether the normal distribution is ap-

plicable or not, appears relevant for practitioners in these fields.

Keywords Bayesian Posterior, Gaussian Approximation, Chi-squared and Binomial Distributions

1 General Notions and Definitions

Bayesian statistics distinguishes the observations x_1, \dots, x_N from the parameter ξ that "conditions" them. Bayes' theorem — given below — expresses the uncertainty about ξ via a probability distribution of ξ conditioned by the observations x_1, \dots, x_N . This so-called posterior distribution becomes always narrower with increasing N . By consequence the "true value" of ξ is approached more and more closely. Simultaneously the posterior approaches a Gaussian distribution. This is a consequence of the fact that the posterior after N observations becomes the N -th power of the posterior from one observation.

The present article, based on Bayesian statistics, derives the minimal N needed for the Gaussian approximation. In the first part, comprising Sects. 2 and 3, the variables x and ξ are defined along the real axis as a whole. A chi-squared distribution provides an example. In the second part, i.e. in Sects. 4 and 5, the parameter ξ is defined on a finite interval of the real axis. The so-called trigonometric distribution serves as example.

We explain some notions used throughout the following text.

A statistical model $p(x|\xi)$ formulates the relation between the observations x and the parameter ξ . A statistical model is

normalised according to

$$\int_{\text{domain } x} dx p(x|\xi) = 1 \quad (1)$$

for every ξ . Here, “*domain x*” means that the integral extends over the domain of definition of the observed x .

Bayes’ theorem [4] states the posterior distribution

$$P(\xi|x) = \frac{p(x|\xi)\mu(\xi)}{m(x)}, \quad (2)$$

of ξ conditioned by x . The Bayesian prior distribution $\mu(\xi)$ serves also as the measure of integration over ξ , see Eq. (3) and the explanation there. The quantity

$$m(x) = \int_{\text{domain } \xi} d\xi p(x|\xi)\mu(\xi) \quad (3)$$

normalises $P(x|\xi)$ to unity. Equation (2) gives the posterior in the case of a single observation x . There is usually more than one observation, but in the present text there shall be only one parameter ξ . For the case of N observations $\mathbf{x} = (x_1, \dots, x_N)$ the posterior is formulated in Sects. 2.4 and 4.4.

The Fisher information [16, 17, 18, 19] is the expectation value

$$F(\xi) = \int_{\text{domain } x} dx p(x|\xi) \left[\frac{\partial}{\partial \xi} \ln p(x|\xi) \right]^2 \quad (4)$$

of the quantity $[\partial/\partial \xi \ln p(x|\xi)]^2$. The Fisher information is always positive. When p is a Gaussian distribution then F is the inverse of the mean square value of the Gaussian.

Form invariance is a symmetry relation between the observation x and the condition ξ . It is defined by a group — in the sense of the theory of Lie groups [24, 32] — of transformations that leave $p(x|\xi)$ unchanged when applied to both, x and ξ . See Chap. 6 of [25] and the textbooks [1, 7].

In the examples of the present text the symmetry shows up by the fact that the model $p(x|\xi)$ depends on the difference $x - \xi$. The symmetry group then consists of all possible simultaneous translations of x and ξ by the same shift. Many statistical models can be reformulated such as to display translational form invariance. In section 2 the details of translational form invariance are discussed. It is taken as the starting point for a Gaussian approximation to the posterior distribution since the Gaussian itself is form invariant under translations. The invariant measure of the group of translations is identified with the prior distribution. Here, this implies

$$\mu(\xi) \equiv \text{const.} \quad (5)$$

In Sects. 2 and 4 two different cases are considered: Continuous x and ξ within a chi-squared model and dichotomic x within the binomial model. Both cases are widespread and of practical interest for many applications.

A likelihood function $\mathcal{L}_N(\xi)$ is proportional to the probability density $p_N(\mathbf{x}|\xi)$ of N observations \mathbf{x} considered as a function of ξ while \mathbf{x} is given. When — in the present context of Eq. (5) — the domain of definition of ξ extends over the whole

real axis, the likelihood function possesses a maximum: Since the posterior is normalised, it must tend to zero when ξ goes to infinity.

The value ξ^{ML} where the maximum occurs, is called the *maximum likelihood* or *ML estimator* of the “true value” of ξ . For every series of observations \mathbf{x} there is a ML estimator $\xi^{\text{ML}} = \xi^{\text{ML}}(\mathbf{x})$. In sections 2.4 and 4.4 one shall see that ξ^{ML} is the *sufficient statistic* [37] of the model — a notion widely discussed in the development of the Rasch model [53, 49, 55, 13, 14, 21, 25].

It might appear that the constancy of the prior is nothing but the “principle of indifference”, well-known in the history of statistical and Bayesian reasoning, well-known also for its difficulties when applied to transform probability densities [11, 20]. Note that in the present paper the constancy of the prior occurs as a consequence of the well-defined group theoretical property of “form invariance” [26].

In Sects. 3.1 and 5 the likelihood function of a model $p(x|\xi)$ is compared to the Gaussian model

$$G(x|\xi) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\xi)^2}{2\sigma^2}\right), \quad -\infty < x, \xi < \infty. \quad (6)$$

The N -fold Gaussian model, i.e. the distribution of N observations $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, is

$$G_N(\mathbf{x}|\xi) = \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \prod_{k=1}^N \exp\left(-\frac{(x_k-\xi)^2}{2\sigma^2}\right). \quad (7)$$

This yields the posterior distribution

$$G_N(\xi|\mathbf{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left(-\frac{N}{2\sigma^2}(\langle x \rangle - \xi)^2\right), \quad (8)$$

where $\langle x \rangle$ is the average

$$\langle x \rangle = \frac{1}{N} \sum_{k=1}^N x_k, \quad (9)$$

see appendix I. The Fisher information of the Gaussian (6) is

$$F^{\text{Gauss}} \equiv \sigma^{-2}. \quad (10)$$

With increasing number N of observations the posterior of any model p with the prior (5) assumes the Gaussian form and becomes always narrower tending towards Dirac’s delta distribution. We shall determine the minimal N which allows to approximate a given posterior by the Gaussian (8) within the interval $|\xi^{\text{ML}} - \xi| < 3\sigma/\sqrt{N}$. This interval contains 99.73 percent of the area under the Gaussian function. Two examples are studied, the chi-squared model in Sect. 3 and the binomial model in Sect. 5. In the two cases the minimal N strongly differ.

2 Form Invariance Along the Real Axis

The present section considers a statistical model $p(x|\xi)$ where both, the observation x and the parameter ξ , are defined

all along the real axis. Furthermore, x and ξ shall be related via form invariance. These two properties allow a general approximation to the logarithmic likelihood function which contains the sum over $\ln p(x_k - \xi)$. These logarithms are N random numbers since the x_k are random, whence the sum over the $\ln p(x_k - \xi)$ will have a Gaussian distribution for sufficiently large N . Yet the central limit theorem does not allow to answer the question, how large N must be for the Gaussian approximation to apply. However, the sum over the $\ln p(x_k - \xi)$ becomes equal to the N -fold (negative) Kullback-Leibler distance $H(\xi^{\text{ML}}|\xi)$, see Ref. [38] and Eq. (24). The Kullback-Leibler distance is a quantity that measures the distance between the distributions $p(x|\xi^{\text{ML}})$ and $p(x|\xi)$. It will be represented by a Taylor expansion with respect to ξ . With increasing N its terms of higher order become negligible as compared to the term of second order. This leads to the criterion for the Gaussian approximation.

2.1 The prior distribution

The above-mentioned two properties of the model $p(x|\xi)$ entail that p depends on the difference $x - \xi$ and only on this difference. Thus the model reads

$$p(x|\xi) = p(x - \xi), \quad -\infty < x, \xi < \infty. \quad (11)$$

Neither Bayes [4] nor Laplace [39] who independently established Bayes' theorem, gave a prescription to obtain the prior distribution. Form invariance gives us the prescription: The prior shall be invariant under the symmetry group of translations [32, 34, 35, 25].

Under the translation

$$\xi' = \xi + a \quad (12)$$

the prior transforms as a density, i.e. $\mu(\xi)$ goes over into

$$\begin{aligned} \mu_T &= \mu(\xi' - a) \left| \frac{d\xi}{d\xi'} \right| \\ &= \mu(\xi' - a). \end{aligned} \quad (13)$$

This shall be independent of a ; thus μ is constant as foreseen in Eq. (5). See also Eq. (33) in Sect. 2.3.

For N events \mathbf{x} , conditioned by one and the same value of ξ , the model is

$$p_N(\mathbf{x}|\xi) = \prod_{k=1}^N p(x_k - \xi), \quad -\infty < x_k, \xi < \infty. \quad (14)$$

The posterior is

$$P_N(\xi|\mathbf{x}) = \mu(\xi) \frac{\prod_{k=1}^N p(x_k - \xi)}{m(\mathbf{x})}, \quad (15)$$

where

$$m(\mathbf{x}) = \int_{-\infty}^{\infty} d\xi \mu(\xi) \prod_{k=1}^N p(x_k - \xi). \quad (16)$$

Since μ is constant its value drops out of the posterior and P becomes

$$P_N(\xi|\mathbf{x}) = \frac{\prod_{k=1}^N p(x_k - \xi)}{\int_{-\infty}^{\infty} d\xi' \prod_{k=1}^N p(x_k - \xi')}. \quad (17)$$

One can numerically calculate this expression, determine the shortest interval in which ξ is found with the probability of 99.73 percent, and thus obtain an error interval for ξ , see Chap. 3 of [25]. In Sect. 2.4 this procedure is replaced by considering the logarithm of the likelihood function. This will show that the posterior P_N of Eq. (17) tends towards a Gaussian with increasing N .

2.2 The Maximum-Likelihood Estimator

The expression $p_N(\mathbf{x}|\xi)$ of Eq. (14) is called a likelihood function $\mathcal{L}_N(\xi)$ when it is considered as a function of ξ , while \mathbf{x} is given.

A likelihood function in the context of translational invariance possesses a maximum since it is a normalisable function defined all along the real axis. When there are several maxima, one must look for the absolute maximum or even redefine the model such that there is an absolute maximum. The place ξ^{ML} , where the maximum occurs, depends on the observed \mathbf{x} . Hence, $\xi^{\text{ML}} = \xi^{\text{ML}}(\mathbf{x})$ is a function of \mathbf{x} . The value ξ^{ML} is called the maximum likelihood (ML) estimator of the parameter ξ . For the example of the chi-squared model in Sect. 3.2 the ML estimator is given in appendix B.

The ML estimator has been introduced by R.A. Fisher [15, 16, 2]. It estimates the "true value" of ξ which conditions the observations \mathbf{x} . For every finite N , however, the true value remains hidden. With $N \rightarrow \infty$ the ML estimator converges to it. An example is given by the Gaussian model (7); the average (9) is its ML estimator.

Section 2.4 shows that $\xi^{\text{ML}}(\mathbf{x})$ is the sufficient statistic of the model p , i.e. the observations \mathbf{x} enter into the posterior distribution only via ξ^{ML} . Compare page 22 of [36] and Sect. 3.1.3 of [21] and chapters 2 and 3 of Ref. [25].

Neyman and Scott [45] have argued against ML estimation. Their argument says that a bias may remain between the expectation value of ξ and the ML estimator. This has caused a considerable debate [61, 56]. The argument of Neyman and Scott is bound to the distinction between between a "structural" and an "incidental" parameter. The model they studied, is form invariant which means that there is a Lie group of transformations that leaves it invariant. Each of the two parameters describes a subgroup. The elements of the different subgroups do not commute with each other. One of the subgroups describes translations, the other one describes dilations, see, e.g., Sect. 7.3 of [25]. The present article describes a more basic situation. We also have two parameters; however, the present subgroups are identical. Both are translational.

At least in the present context of translational form invariance such a bias becomes arbitrarily small with increasing N — as is shown for the chi-squared model of Eq. (45) in Sect. 3.2. It possesses such a bias; but the fact that the posterior tends to a Gaussian with increasing N implies that the bias goes to

zero. Future research should show whether this holds also for the model discussed by Neyman and Scott.

We study the transition of the posterior to a Gaussian distribution by help of the logarithm of P_N given in Eq. (15). Since both, $\mu(\xi)$ and $m(\mathbf{x})$, are independent of ξ , this likelihood function is proportional to the product of the $p(x_k - \xi)$, i.e.

$$\mathcal{L}_N(\xi) \propto \prod_{k=1}^N p(x_k - \xi). \quad (18)$$

Therefore the logarithm of \mathcal{L}_N is — up to an additive constant — the sum over the logarithms $\ln p(x_k - \xi)$,

$$\text{const} + \ln \mathcal{L}_N(\xi) = \sum_{k=1}^N \ln p(x_k - \xi). \quad (19)$$

For sufficiently large N the sum over the logarithms is expressed by the expectation value of $\ln p(x - \xi)$ which in principle requires

$$\text{const} + \ln \mathcal{L}_N(\xi) = N \int_{-\infty}^{\infty} dx p(x - \xi^{\text{true}}) \ln p(x - \xi). \quad (20)$$

This integral is the expectation value of $\ln p(x - \xi)$ taken with the distribution conditioned by the “true value” of ξ . Jaynes and Bretthorst [31] have called it the “asymptotic” likelihood function, asymptotic in the sense of $N \rightarrow \infty$. The true value of ξ remains, however, hidden for every finite N . We replace it by the ML estimator ξ^{ML} obtained from the observations $\mathbf{x} = (x_1, \dots, x_N)$. Then the “asymptotic” likelihood function becomes

$$\text{const} + \ln \mathcal{L}_N(\xi) = N \int_{-\infty}^{\infty} dx p(x - \xi^{\text{ML}}) \ln p(x - \xi). \quad (21)$$

We want to define const such that $\ln \mathcal{L}_N$ becomes a Kullback-Leibler distance. This is reached when const is set to

$$\text{const} = -N \int_{-\infty}^{\infty} dx p(x - \xi^{\text{ML}}) \ln p(x - \xi^{\text{ML}}) \quad (22)$$

Then the “asymptotic” likelihood function becomes

$$\ln \mathcal{L}_N(\xi) = N \int_{\text{domain } x} dx p(x - \xi^{\text{ML}}) \ln \frac{p(x - \xi)}{p(x - \xi^{\text{ML}})}. \quad (23)$$

We call this integral the functional

$$H(\xi^{\text{ML}}|\xi) = \int_{\text{domain } x} dx p(x - \xi^{\text{ML}}) \ln \frac{p(x - \xi)}{p(x - \xi^{\text{ML}})}. \quad (24)$$

It is the negative of the Kullback-Leibler distance [38, 27, 9] between the distributions conditioned by ξ^{ML} and by ξ . Thus the logarithmic likelihood function becomes

$$\ln \mathcal{L}_N(\xi) = N H(\xi^{\text{ML}}|\xi). \quad (25)$$

The Taylor expansion of H with respect to ξ will yield our criterion by requiring that terms of higher than the second order be negligible.

How large must N be for $\ln \mathcal{L}_N$ to become equal to the expectation value (23)? We follow the assumption that there is

such an N . For this and larger N equation (23) holds: The logarithmic likelihood function of the Gaussian (8) is

$$\ln \mathcal{L}_N^{\text{Gauss}}(\xi) = N \left[-\frac{(\langle x \rangle - \xi)^2}{2\sigma^2} \right], \quad (26)$$

where the quantity in rectangular brackets is the functional (24) for the Gaussian (6),

$$H^{\text{Gauss}}(\xi^{\text{ML}}|\xi) = -\frac{(\langle x \rangle - \xi)^2}{2\sigma^2}. \quad (27)$$

The quantity $\langle x \rangle$, given in Eq. (9), is the ML estimator of the Gaussian model. One sees that Eqs. (21) and (25) are fulfilled by the Gaussian model.

Equation (103) in appendix A shows that the maximum value of $H(\xi^{\text{ML}}|\xi)$ — and thus of the logarithmic likelihood — occurs at $\xi = \xi^{\text{ML}}$ for every distribution of the form $p(x - \xi)$.

According to Eq. (25) the likelihood $\mathcal{L}_N(\xi)$ has the form of the N -th power of $\mathcal{L}_{N=1}(\xi)$ since $\ln \mathcal{L}_N(\xi)$ is proportional to N . The translational invariance of p given by Eq. (11) leads to a translational invariance of H since

$$\begin{aligned} H(\xi^{\text{ML}}|\xi) &= \int_{-\infty}^{\infty} dx p(x - \xi^{\text{ML}}) \ln \frac{p(x - \xi)}{p(x - \xi^{\text{ML}})} \\ &= \int_{-\infty}^{\infty} dx' p(x' - \xi^{\text{ML}} + \xi) \ln \frac{p(x')}{p(x' - \xi^{\text{ML}} + \xi)} \\ &= H(\xi^{\text{ML}} - \xi|0). \end{aligned} \quad (28)$$

This result has been obtained by substituting

$$x' = x - \xi. \quad (29)$$

Thus $H(\xi^{\text{ML}}|\xi)$ depends on the difference $\xi^{\text{ML}} - \xi$ and only on this difference.

When there are isolated points where $p(x')$ or $p(x' - \xi^{\text{ML}} + \xi)$ vanish, then appendix H shows that the integral exists despite the divergence of the integrand in (28).

2.3 The Fisher Information

The information that carries the name of R.A. Fisher [17] is a central concept of statistics and estimation theory. It provides the prior, if form invariance holds. The Fisher information F is defined as the expectation value of the squared derivative of the logarithmic likelihood function, see Refs. [16, 17, 18, 19] and Sect. 4.2 of [21], i.e.

$$\begin{aligned} F(\xi) &= \int_{\text{domain } x} dx p(x|\xi) \left[\frac{\partial}{\partial \xi} \ln p(x|\xi) \right]^2 \\ &= - \int_{\text{domain } x} dx p(x|\xi) \frac{\partial^2}{\partial \xi^2} \ln p(x|\xi). \end{aligned} \quad (30)$$

The two expressions equal each other because p is normalised to unity for every ξ , see appendix C. Note that the definition (30) refers to a single observation, see [42]. Other authors, however, define the Fisher information for multiple observations [47]. The integration in (30) extends over the full domain

of definition of x . Given the domain of Eq. (11) one has by use of the substitution (29)

$$\begin{aligned} F &= - \int_{-\infty}^{\infty} dx p(x - \xi) \frac{\partial^2}{\partial \xi^2} \ln p(x - \xi) \\ &= - \int_{-\infty}^{\infty} dx' p(x') \frac{d^2}{dx'^2} \ln p(x'). \end{aligned} \quad (31)$$

Thus the Fisher information of the model (11) is independent of ξ . This is a consequence of form invariance under translations.

By the definitions of H in (24) and F in (30) one recognises that

$$F = - \left. \frac{\partial^2}{\partial \xi^2} H(\xi^{\text{ML}}|\xi) \right|_{\xi=\xi^{\text{ML}}}. \quad (32)$$

Since F is positive, the second derivative of H is seen to be negative at $\xi = \xi^{\text{ML}}$. This means that the second derivative of $\mathcal{L}(\xi)$ is negative, as it should be at a maximum.

The prior distribution μ is proportional to the square root of F ,

$$\mu \propto \sqrt{F}, \quad (33)$$

according to Hartigan [26] and Jaynes [30]. A measure is the inverse unit of length on the scale of ξ .

By identifying the integration measure on the scale of ξ with the Bayesian prior distribution, we follow Jeffreys [32] as was done by other authors [33, 29, 35, 48]. The prior is independent of N . Since it is independent of ξ and ξ^{ML} , a given difference $\xi^{\text{ML}} - \xi$ describes the same length of way everywhere on the scale of ξ . See also [3, 5, 62, 63, 25] and the footnote ¹.

2.4 The Posterior of a Form Invariant Model

From Eq. (25) and the fact that the prior is constant follows that the posterior distribution is

$$P_N(\xi|\mathbf{x}) = \mathcal{N}_N \exp(NH(\xi^{\text{ML}}|\xi)), \quad (34)$$

where \mathcal{N}_N normalises P_N to unity. The posterior from N observations has the form of the N -th power of the posterior for $N = 1$. Therefore with increasing N the posterior is more and more restricted to the immediate neighborhood of the maximum at $\xi = \xi^{\text{ML}}$. At this maximum $H(\xi^{\text{ML}}|\xi^{\text{ML}})$ vanishes, since H equals the Kullback-Leibler distance between the distributions $p(x - \xi^{\text{ML}})$ and $p(x - \xi)$. Thus for $\xi = \xi^{\text{ML}}$ the exponential in (34) has the value of unity,

$$\exp(NH(\xi^{\text{ML}}|\xi^{\text{ML}})) = 1. \quad (35)$$

With increasing N the curvature of the likelihood function $\exp(NH(\xi^{\text{ML}}|\xi))$ at its maximum increases according to NF , were F is the Fisher information at $N = 1$. The Fisher information is independent of ξ in the present context. Thus the likelihood function behaves as the Gaussian (8), see Eqs.

(26) and (27). It does so in a suitable interval around the maximum value (35) of the likelihood. Thus within that interval the likelihood becomes a Gaussian function.

One also sees that the posterior depends on the observations \mathbf{x} via and only via the estimator $\xi^{\text{ML}} = \xi^{\text{ML}}(\mathbf{x})$; this means that $\xi^{\text{ML}}(\mathbf{x})$ is the sufficient statistic [36]. The notion of sufficient statistic has been widely discussed in the development of the Rasch model of item response theory [53, 49, 13, 14, 55, 21, 25]. We shall come to it in Sect. 4.5 where the binomial model is treated.

3 The Criterion for the Gaussian Approximation

A criterion will now be formulated that allows to find the minimal N so that the Gaussian distribution (8) can be accepted as approximating the posterior P_N in Eq. (34).

The Gaussian approximation is justified whenever the Taylor expansion of $\ln \mathcal{L}_N(\xi)$ up to the second order in ξ is “sufficiently precise”. This is the expansion of $H(\xi^{\text{ML}}|\xi)$ with respect to ξ at the point $\xi = \xi^{\text{ML}}$. In Sect. 3.1 the Taylor expansion is written down and our understanding of “sufficiently precise” is defined. This yields the desired criterion. In Sect. 3.2, the criterion is applied to a chi-squared model.

3.1 Formulation of the Criterion

The Taylor expansion of H up to an arbitrary order n is

$$H(\xi^{\text{ML}}|\xi) = \sum_{\nu=1}^n \frac{(\xi - \xi^{\text{ML}})^\nu}{\nu!} \left[\frac{\partial^\nu}{\partial \xi^\nu} H(\xi^{\text{ML}}|\xi) \right] \Big|_{\xi=\xi^{\text{ML}}} + R. \quad (36)$$

The quantity R is the remainder. The zero-th order, $\nu = 0$, of this expansion vanishes. We expand up to $n = 2$ and choose Lagrange’s remainder among several versions suggested in Sect. 0.317 of [22]. This is

$$R = \frac{(\xi - \xi^{\text{ML}})^3}{3!} \frac{\partial^3}{\partial \xi^3} H(\xi^{\text{ML}}|\xi') \Big|_{\xi'=\xi^{\text{ML}}+(\xi-\xi^{\text{ML}})\Theta}, \quad (37)$$

$0 < \Theta < 1$.

The value of Θ remains open except for the fact that it lies between 0 and 1.

After N observations the Gaussian (8) is accepted as a valid approximation if the remainder R is negligible for all $\xi - \xi^{\text{ML}}$ in the interval

$$-3\sigma \frac{1}{\sqrt{N}} < \xi - \xi^{\text{ML}} < 3\sigma \frac{1}{\sqrt{N}}. \quad (38)$$

containing 99.73 percent of the area of the Gaussian function. We call it the 3σ -interval of the Gaussian approximation (8), See FIG. 1. The value of σ is given by the Fisher information

$$F = \sigma^{-2} \quad (39)$$

of the model p . This value is the expected curvature of the likelihood function. The Fisher information together with the measure μ has been defined in Sect. 2.3.

¹The prior distribution by Zellner [62] is constructed by help of Shannon’s information. For the model (11) Zellner agrees with the measures by Jeffreys and Kass in that Zellner’s prior is constant. However, his prior does not behave as a density under transformations. So it is not generally proportional to the square root of (32).

The term of $\nu = 0$ in the expansion (36) vanishes according to the definition (24) of the functional H . The first derivative $\frac{\partial}{\partial \xi} H$ vanishes at $\xi = \xi^{\text{ML}}$ because H attains its maximum value for $\xi = \xi^{\text{ML}}$. The second derivative $\frac{\partial^2}{\partial \xi^2} H$ at $\xi = \xi^{\text{ML}}$ yields the negative Fisher information according to Eq. (32). The third derivative — in the remainder — is calculated by help of the translational invariance (28) of H . This gives

$$\begin{aligned} & \frac{\partial^3}{\partial \xi'^3} H(\xi^{\text{ML}}|\xi') \Big|_{\xi'=\xi^{\text{ML}}+(\xi-\xi^{\text{ML}})\Theta} \\ &= \int_{-\infty}^{\infty} dx p(x - \xi^{\text{ML}}) \frac{\partial^3}{\partial \xi'^3} \ln p(x - \xi') \Big|_{\dots} \\ &= - \int_{-\infty}^{\infty} dx \frac{\partial^3}{\partial \xi'^3} p(x - \xi^{\text{ML}} + \xi') \ln p(x) \Big|_{\dots} \end{aligned} \quad (40)$$

The subscripts $|\dots$ in the r.h.s. of this equation mean repetitions of the subscript on the l.h.s.

In order to decide whether the remainder R is small enough to be neglected, we must find below the largest absolute value $|H^{(3)}|_{\text{max}}$ of (40). The largest value of $(\xi - \xi^{\text{ML}})^3$ occurs at the upper end of the interval (38), whence the maximum absolute value of the remainder is

$$|R|_{\text{max}} = \frac{1}{3!} \left(\frac{3\sigma}{\sqrt{N}} \right)^3 |H^{(3)}|_{\text{max}}. \quad (41)$$

This shall be small compared to the second-order-term

$$\frac{1}{2} \left(\frac{3\sigma}{\sqrt{N}} \right)^2 F$$

everywhere in the interval (38). Thus we require

$$\frac{1}{2} \left(\frac{3\sigma}{\sqrt{N}} \right)^2 F \gg \frac{1}{6} \left(\frac{3\sigma}{\sqrt{N}} \right)^3 |H^{(3)}|_{\text{max}}. \quad (42)$$

This gives, by use of (39),

$$1 \gg \frac{|H^{(3)}|_{\text{max}}}{N^{1/2} F^{3/2}}. \quad (43)$$

We consider this condition to be fulfilled if

$$0.1 \geq \frac{|H^{(3)}|_{\text{max}}}{N^{1/2} F^{3/2}}. \quad (44)$$

The last inequality is our criterion for the validity of the Gaussian approximation. Replacing the requirement (43) by (44), we follow a convention. A step of one “order of magnitude” is usually considered to realise the requirement that one object be large compared to another one.

3.2 The Chi-Squared Model

As an example let us consider the model

$$p(x - \xi) = \exp(x - \xi - e^{x-\xi}), \quad -\infty < x, \xi < \infty. \quad (45)$$

It is normalised to unity, see section 8.312, no. 10 of [22]. This is a chi-squared distribution with two degrees of freedom

transformed such that it depends on the difference $x - \xi$ between event and parameter, see Eq. (114) in appendix D. The ML estimator is $\xi^{\text{ML}} = x$, see Eq. (106) in appendix B.

In the expansion (36) the terms of order $\nu = 0, 1$ vanish according to Sect. 3.1. The derivatives of H are obtained as follows. The first line of (28) with (45) gives

$$\begin{aligned} & \frac{\partial}{\partial \xi} H(\xi^{\text{ML}}|\xi) \\ &= \frac{\partial}{\partial \xi} \int_{-\infty}^{\infty} dx \left(\exp(x - \xi^{\text{ML}} - e^{x-\xi^{\text{ML}}}) \right) \\ & \quad \left[\xi^{\text{ML}} - \xi - e^{x-\xi} + e^{x-\xi^{\text{ML}}} \right] \\ &= \int_{-\infty}^{\infty} dx \left(\exp(x - \xi^{\text{ML}} - e^{x-\xi^{\text{ML}}}) \right) \\ & \quad \left[-1 + e^{x-\xi} \right] \\ &= - \int_{-\infty}^{\infty} dx \exp(x - \xi^{\text{ML}} - e^{x-\xi^{\text{ML}}}) \\ & \quad + \int_{-\infty}^{\infty} dx \exp(x - \xi^{\text{ML}} - e^{x-\xi^{\text{ML}}} + x - \xi). \end{aligned} \quad (46)$$

In the last line the first term on the r.h.s. has the value -1 due to the normalisation of the model 45. Whence, we obtain

$$\frac{\partial}{\partial \xi} H(\xi^{\text{ML}}|\xi) = -1 + \int_{-\infty}^{\infty} dx \left(\exp(x - \xi^{\text{ML}} - e^{x-\xi^{\text{ML}}} + x - \xi) \right). \quad (47)$$

The substitution

$$x' = x - \xi^{\text{ML}} \quad (48)$$

yields

$$\frac{\partial}{\partial \xi} H(\xi^{\text{ML}}|\xi) = -1 + \int_{-\infty}^{\infty} dx' \left(\exp(2x' - e^{x'} - \xi + \xi^{\text{ML}}) \right). \quad (49)$$

For $\xi = \xi^{\text{ML}}$ this becomes

$$\begin{aligned} \frac{\partial}{\partial \xi} H(\xi^{\text{ML}}|\xi) \Big|_{\xi=\xi^{\text{ML}}} &= -1 + \int_{-\infty}^{\infty} dx' \exp(2x' - e^{x'}) \\ &= -1 + \Gamma(2), \end{aligned} \quad (50)$$

see appendix E. This yields

$$\begin{aligned} \frac{\partial}{\partial \xi} H(\xi^{\text{ML}}|\xi) \Big|_{\xi=\xi^{\text{ML}}} &= -1 + 1 \\ &= 0. \end{aligned} \quad (51)$$

This result was expected since $H(\xi^{\text{ML}}|\xi)$ has an extreme value at $\xi = \xi^{\text{ML}}$.

The second derivative of H with respect to ξ is

$$\begin{aligned} \frac{\partial^2}{\partial \xi^2} H(\xi^{\text{ML}}|\xi) &= - \int_{-\infty}^{\infty} dx' \exp(2x' - e^{x'}) e^{-\xi + \xi^{\text{ML}}} \\ &= -e^{\xi^{\text{ML}} - \xi} \int_{-\infty}^{\infty} dx' \exp(2x' - e^{x'}) \\ &= -e^{\xi^{\text{ML}} - \xi} \Gamma(2) \\ &= -e^{\xi^{\text{ML}} - \xi}. \end{aligned} \quad (52)$$

The definition of $H(\xi^{ML}|\xi)$ in the first line of Eq. (28) together with the substitution (48) yields a symmetry relation of the functional H ,

$$\begin{aligned} H(\xi^{ML}|\xi) &= \int_{-\infty}^{\infty} dx' \exp(x' - e^{x'}) \\ &\ln \frac{\exp(x' - e^{x'+\xi^{ML}-\xi} + \xi^{ML} - \xi)}{\exp(x' - e^{x'})} \\ &= H(0|\xi - \xi^{ML}). \end{aligned} \tag{53}$$

The last line of Eq. (52) yields the derivatives

$$\frac{\partial^\nu}{\partial \xi^\nu} H(\xi^{ML}|\xi) = (-1)^{\nu+1} e^{\xi^{ML}-\xi}, \quad \nu \geq 2. \tag{54}$$

One especially obtains the Fisher information

$$\begin{aligned} F &= - \left. \frac{\partial^2}{\partial \xi^2} H(0|\xi - \xi^{ML}) \right|_{\xi=\xi^{ML}} \\ &= 1. \end{aligned} \tag{55}$$

Equation (54) gives the maximum $|H^{(3)}|_{\max}$ of the absolute value of (40) in the interval (38) to be

$$\begin{aligned} |H^{(3)}|_{\max} &= \exp\left(3 \frac{\sigma}{N^{1/2}}\right) \\ &= \exp\left(\frac{3}{F^{1/2} N^{1/2}}\right) \\ &= \exp\left(\frac{3}{N^{1/2}}\right). \end{aligned} \tag{56}$$

Then Eq. (44) yields the criterion

$$0.1 \geq \frac{1}{N^{1/2}} \exp\left(\frac{3}{N^{1/2}}\right). \tag{57}$$

Table 1 shows that the inequality (57) is satisfied if

$$N \geq 160. \tag{58}$$

Thus the posterior of the chi-squared distribution (45) with two degrees of freedom can be considered Gaussian when the number N of events is larger than or equal to 160. This shows that an intuition based on the central limit theorem, were a current ‘‘rule of thumb’’ says $N \approx 30$, can be misleading. See Sect. 7.4.2 of [54] or Hogg et al. [28].

To find the necessary N , we have not used the article by Berry [6] on the accuracy of the Gaussian approximation because Berry does not discuss the measure on the scales of the variables.

To illustrate the approximation of the likelihood functions, see FIG. 1:

In FIG. 1 we compare the logarithmic likelihood function of a Chi-squared model to that of a Gaussian (i.e. a parabola). The former is taken from Eq. (25) together with (49) and $N = 1$; the latter taken from (27) with $\sigma^2 = F = 1$. Note the considerable skewness of the log-likelihood for the Chi-squared model.

Table I

N	r.h.s. of (57)
3	3.263
4	2.241
5	1.711
10	0.817
20	0.437
30	0.316
40	0.254
50	0.216
75	0.163
100	0.135
150	0.104
155	0.102
160	0.100
165	0.098

Table 1. Right hand side of inequality (57) for various values of N .

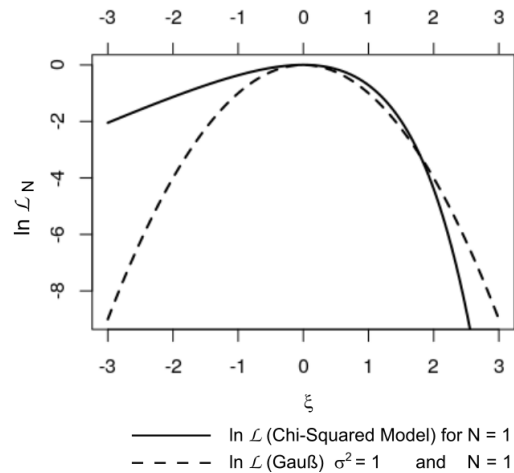


Figure 1. Comparison of the logarithmic likelihood function of a Chi-squared model to that of a Gaussian (i.e. a parabola) for $N = 1$.

4 The Binomial Model

In the following Sects. 4 and 5 the *binomial model* $q(x|\xi)$ is considered. It is also called the model of the ‘‘simple alternative’’ since its event x is restricted to two possible values $x = 0$ or $x = 1$ while ξ is a continuous variable which conditions the probability to find x . By consequence, form invariance cannot be expressed by the difference between x and ξ . Yet form invariance under translations is found by reformulating the model such that it depends on the difference between the ML estimator ξ^{ML} and the parameter ξ , and only on this difference.

4.1 Definition of the Binomial Model

The binomial model is given by

$$q(x|\xi) = [R(\xi)]^x [1 - R(\xi)]^{1-x}, \quad x = 0, 1. \tag{59}$$

Here, $x = 1$ can be interpreted as the “correct” and $x = 0$ as the “false” answer to a given question. Instead of correct or false, the two possibilities may also mean one or the other side of a thrown coin. The function R of the continuous variable ξ is called item response function.

The name of *item response function* (IRF) is due to authors who discussed the ideas of “item response theory” [43, 12, 41, 44, 53, 50, 51, 52] applied in intelligence tests as well as educational tests, e.g. the PISA tests [46]. Differing from them we define the IRF by the requirement that the measure μ on the scale of ξ should be constant. This means that the Fisher information F shall be independent of ξ , see Eq. (33). The Fisher information of the binomial model is defined much as in Eq. (4), except for the integral over x which becomes a sum over $x = 1, 2$. We make use of the second version of Eq. (30) and obtain a differential equation with the solution

$$R(\xi) = \cos^2 \xi, \quad -\pi/2 \leq \xi \leq \pi/2, \quad (60)$$

see appendix J. The domain of definition of the parameter ξ makes sure that the likelihood function possesses a maximum and vanishes towards the ends of the domain. This is analogous to the model $p(x - \xi)$ introduced in Sect. 2.1. The binomial model with the IRF (60) is called the *trigonometric model*.

By the definition of the IRF (60) the Fisher information of the model (59) becomes

$$F \equiv 4, \quad (61)$$

see appendix J; whence the measure (33) on the scale of ξ is

$$\mu \equiv 2. \quad (62)$$

G. Rasch did not consider the measure on the scale of the IRF, he derived the IRF by requiring the property of “specific objectivity” to any measurement [53, 50, 51, 52]. By this property the score of correct answers became the sufficient statistic in his model. The trigonometric IRF, however, is compatible with specific objectivity, see Sect. 3.1 of [21].

For N events $\mathbf{x} = (x_1, \dots, x_N)$ conditioned by one and the same ξ , the binomial model reads

$$q_N(\mathbf{x}\xi) = \prod_{k=1}^N [R(\xi)]^{x_k} [1 - R(\xi)]^{1-x_k}, \quad x_k = 0, 1. \quad (63)$$

This can be rewritten by help of the score s_c of the answers that yield $x = 1$ as well as the number $N - s_c$ of the answers that yield $x = 0$. One obtains

$$q_N(s_c|\xi) = \binom{N}{s_c} [R(\xi)]^{s_c} [1 - R(\xi)]^{N-s_c}, \quad (64)$$

$$0 \leq s_c \leq N, \quad N \geq 1.$$

The quantity $\binom{N}{s_c}$ is a binomial coefficient. By applying the binomial formula one finds that q_N is normalised,

$$\sum_{s_c=0}^N q_N(s_c|\xi) = 1. \quad (65)$$

The posterior distribution will be obtained in Sect. 4.5.

4.2 The ML Estimator for the Trigonometric Model

Given the event s_c in the framework of the distribution (64), the ML estimator is found by solving the the ML equation

$$\frac{\partial}{\partial \xi} \ln q_N(s_c|\xi) = 0. \quad (66)$$

This leads to

$$0 = \frac{\partial}{\partial \xi} [s_c R(\xi) + (N - s_c) \ln(1 - R(\xi))]$$

$$= s_c \frac{R'}{R} - (N - s_c) \frac{R'}{1 - R}. \quad (67)$$

With the IRF (60) this ML-equation becomes

$$0 = -2 \cos \xi \sin \xi \left[\frac{s_c}{\cos^2 \xi} - \frac{N - s_c}{\sin^2 \xi} \right]$$

$$= -2 \frac{s_c - N \cos^2 \xi}{\cos \xi \sin \xi} \quad (68)$$

which is solved by $\xi^{\text{ML}}(s_c)$ such that

$$\cos^2 \xi^{\text{ML}} = \frac{s_c}{N}. \quad (69)$$

The denominator of the r.h.s. of (68) seems to exclude the values $\xi^{\text{ML}} = \pm\pi/2$ and $\xi^{\text{ML}} = 0$ because it vanishes there. These values correspond to the “uniform answers” $s_c = 0$ and $s_c = N$, respectively. We show that the uniform answers are not excluded. In the case of $s_c = 0$ equation (68) turns into

$$0 = \frac{\cos \xi}{\sin \xi} \quad (70)$$

and has the solutions $\xi^{\text{ML}} = \pm\pi/2$. This is conform with (69). In the case of $s_c = N$ equation (68) becomes

$$0 = \frac{1 - \cos^2 \xi}{\cos \xi \sin \xi}$$

$$= \frac{\sin \xi}{\cos \xi} \quad (71)$$

which is solved by $\xi^{\text{ML}} = 0$ and again conforms with Eq. (69).

4.3 The Likelihood Function of the Trigonometric Model in the case of $N = 1$

The likelihood function of the trigonometric model $q_{N=1}$ with R according to (60) is form invariant. To show this we study again the two cases of $s_c = 0, 1$.

(i) Let $s_c = 0$ be observed. Then the likelihood function is proportional to $1 - R(\xi)$ and the ML estimator is $\xi^{\text{ML}} = \pm\pi/2$. Thus the likelihood function is

$$\mathcal{L}_{N=1}(\xi) \propto \sin^2 \xi, \quad -\pi/2 < \xi < \pi/2. \quad (72)$$

This can also be expressed as

$$\mathcal{L}_{N=1}(\xi) \propto \cos^2(\xi - \xi^{\text{ML}}). \quad (73)$$

(ii) Now let $s = 1$ be observed. Then the likelihood function is proportional to $R(\xi)$ whence the ML estimator is $\xi^{ML} = 0$, according to the case of Eq. (71). Thus the likelihood function is

$$\mathcal{L}_{N=1}(\xi) \propto \cos^2 \xi. \tag{74}$$

Since now $\xi^{ML} = 0$ this is again expressed by Eq. (73).

Thus if $N = 1$ and R given by (60) the likelihood function of the trigonometric model depends on the difference $\xi - \xi^{ML}$ and only on this difference. This is what we call form invariance under translations. In summary: The posterior of the trigonometric model q is obtained with a constant prior and it is form invariant for $N = 1$.

The result (73) is found for every value of ξ^{ML} given by Eq. (69). Thus it remains true for any number N of observations. The posterior of the trigonometric model is form invariant under translations in the sense that it depends on the difference $\xi - \xi^{ML}$. However, the variables ξ and ξ^{ML} do not “make the same use” of the interval $-\pi/2 < \xi < \pi/2$ given in Eq. (72): The variable ξ is defined everywhere on this interval. The ML estimator ξ^{ML} assumes only a finite number of values within that interval. In the following Sect. 4.4 the trigonometric model itself — not only its posterior — will be brought into translational form invariance.

4.4 The Trigonometric Model with Translational Invariance

For arbitrary N the ML estimator given by Eq. (69) is not restricted to two values. For sufficiently large N any value in the interval $-\pi/2 \leq \xi^{ML} \leq 0$ can be approached arbitrarily closely. Is it possible to define the binomial model such that the model itself is translationally form invariant? Yes, this is possible and leads to the same Fisher information (61).

Consider the model

$$t(s|\xi) = \mathcal{N} \cos^2(s - \xi), \quad -\pi/2 \leq s, \xi \leq \pi/2, \tag{75}$$

which depends on the difference between the observation s and the condition ξ .

According to the argument Sect. 3.2 on the posterior of the trigonometric model, the ML estimator comes arbitrarily close to every value in the interval $-\pi/2 \leq \xi^{ML} \leq \pi/2$ since, for arbitrary N , it assumes every rational number in this interval. Thus the trigonometric model can be extended to the model 75. The fact that ξ^{ML} is defined within, $-\pi/2 \leq \xi^{ML} \leq \pi/2$ means that ξ is defined in the same interval.

We shall show that (75) is translationally form invariant and has the Fisher information (61) in agreement with the trigonometric model (59) with (60).

The normalisation \mathcal{N} in (75) equals the integral over the domain of s ; it is independent of ξ because the domain given in (75) is an interval of length π which is equal to one period of the periodic function $\cos^2(x - \xi)$. Thus \mathcal{N} is given by the

integral

$$\begin{aligned} \mathcal{N}^{-1} &= \int_{-\pi/2}^{\pi/2} ds \cos^2(s - \xi) \\ &= \int_{-\pi/2}^{\pi/2} ds \cos^2 s. \end{aligned} \tag{76}$$

One obtains

$$\mathcal{N} = 2/\pi, \tag{77}$$

see appendix G.

The ML estimator ξ^{ML} for a given event s occurs at

$$\xi^{ML} = s \tag{78}$$

since the likelihood function $t(s|\xi)$ of (75) becomes maximal when the argument of the \cos^2 -function is zero.

The posterior of $t(s|\xi)$ shall be a function of $\xi^{ML} - \xi$ as it is in the context of the model (45) in Sect. 3. This requires to shift — within the posterior — the value of ξ^{ML} to the value of zero. We can do so since the measure on the scale of ξ is constant. The shift avoids that the distance $|\xi^{ML} - \xi|$ becomes larger than the length π of the interval in which ξ is defined. In the context of the model (45) such a shift was not needed since every difference $\xi^{ML} - \xi$ was contained in the infinite domain $-\infty < \xi < \infty$, where ξ was defined.

Since ξ^{ML} is the sufficient statistic of the model (75) the score s_c of correct answers determines the sufficient statistic when N is given. In this sense the model (75) confirms the requirement of G. Rasch [53, 49, 55, 13, 14] that the score should be the sufficient statistic of the binomial model. See especially Chap. 4.6 of [21] and Chap. 12.3.1 of [25], were so-called Guttman schemes [23] are analysed.

The Fisher information of the model (75) is

$$F(\xi) = -\mathcal{N} \int_{-\pi/2}^{\pi/2} ds \cos^2(s - \xi) \frac{\partial^2}{\partial \xi^2} \ln \cos^2(s - \xi) \tag{79}$$

by Eq. (30). For the second derivative in the integrand one finds

$$\frac{\partial^2}{\partial \xi^2} \ln \cos^2(s - \xi) = -\frac{2}{\cos^2(s - \xi)}. \tag{80}$$

Together with (77) this gives

$$F(\xi) \equiv 4. \tag{81}$$

Thus the Fisher information of the translationally invariant model (75) agrees with the Fisher information (61) obtained from the binomial model (59).

4.5 The Functional H for the Trigonometric Model with Translational Invariance

We express the logarithmic likelihood function $\ln \mathcal{L}_N$ for N events given by the model (75) in analogy to Eq. (25) as

$$\begin{aligned} \ln \mathcal{L}_N &= N H(\xi^{ML}|\xi) \\ &= N \int_{-\pi/2}^{\pi/2} ds t(s - \xi^{ML}) \ln \frac{t(s - \xi)}{t(s - \xi^{ML})}. \end{aligned} \tag{82}$$

The integration extends over one period of the \cos^2 -function. Inserting (75) into (82) one obtains

$$\ln \mathcal{L}_N = N \mathcal{N} \int_{-\pi/2}^{\pi/2} ds \cos^2(s - \xi^{\text{ML}}) \ln \frac{\cos^2(s - \xi)}{\cos^2(s - \xi^{\text{ML}})}. \quad (83)$$

The functional $H(\xi^{\text{ML}}|\xi)$ exists and has the property (28) because the integral in (83) exists although the function $\cos^2 s'$ with, e.g.,

$$s' = s - \xi \quad (84)$$

will vanish at an isolated point s'_0 and $\ln \cos^2 s'_0$ will diverge. Appendix H shows that the integral over an ϵ -interval that includes s'_0 , does exist. Here, ϵ may be arbitrarily small. The value of this integral is not proportional to ϵ ; instead it is proportional to $\epsilon \ln \epsilon$. Although this is not as small as ϵ , it approaches zero when $\epsilon \rightarrow 0$. Thus the logarithmic likelihood $\ln \mathcal{L}_N$ of Eq. (83) exists even if the integration runs over a point where $\ln \cos^2$ diverges.

Applying the substitution (84) to the integrand in (83) one obtains

$$\begin{aligned} \ln \mathcal{L}_N &= N \mathcal{N} \int_{-\pi/2}^{\pi/2} ds' \cos^2 s' \ln \frac{\cos^2(s' - \xi + \xi^{\text{ML}})}{\cos^2 s'} \\ &= NH(0|\xi - \xi^{\text{ML}}). \end{aligned} \quad (85)$$

The substitution does not require a shift of the limits of integration because the integrand is periodic with a period of length π and the integration covers one period. Thus the functional $H(\xi^{\text{ML}}|\xi)$ for the translationally invariant trigonometric model has the property (28) which was found earlier for models defined on the entire real axis.

The fact that \mathcal{L}_N equals N times $H(0|-\alpha)$ shows that Eq. (34) holds here, too: The posterior distribution from N events of the trigonometric model is proportional to the N -th power of the posterior from one event.

5 The Gaussian Approximation to the Trigonometric Model with Translational Invariance

The present section establishes the criterion for the validity of the Gaussian approximation to the likelihood function of the trigonometric model with translational invariance. In analogy to the procedure in Sect. 3 we expand $H(\xi^{\text{ML}}|\xi)$ into a Taylor series with respect to ξ at ξ^{ML} . Again the zero-th order of this expansion vanishes by the definition of H in Eq. (24). Again the likelihood function is related to H via Eq. (25). Using the abbreviation (120), introduced in appendix F, we obtain

$$H(0|\xi - \xi^{\text{ML}}) = H(\xi^{\text{ML}} - \xi|0). \quad (86)$$

Appendix F furthermore shows that the expression (86) is mirror symmetric with respect to $\xi^{\text{ML}} - \xi = 0$, i.e.

$$H(0|\xi - \xi^{\text{ML}}) = H(0|\xi^{\text{ML}} - \xi). \quad (87)$$

Therefore all odd derivatives of $H(0|\xi^{\text{ML}} - \xi)$ vanish at $\xi^{\text{ML}} - \xi = 0$, thus

$$\frac{\partial^{2\nu+1}}{\partial \xi^{2\nu+1}} H(0|\xi^{\text{ML}} - \xi) \Big|_{\xi=\xi^{\text{ML}}} = 0, \quad \nu = 0, 1, 2, \dots \quad (88)$$

By consequence, the remainder of the present expansion is not of the third order as it is in Sect. 3.1; it rather is of the fourth order.

All even derivatives $\frac{\partial^{2\nu}}{\partial \alpha^{2\nu}} H(0|\alpha)$ exist. Therefore in the present context the Taylor expansion (36) becomes

$$\begin{aligned} H(\xi^{\text{ML}}|\xi) &= -\frac{(\xi^{\text{ML}} - \xi)^2}{2} F + \\ &\frac{(\xi^{\text{ML}} - \xi)^4}{4!} \frac{\partial^4}{\partial \xi^4} H(\xi^{\text{ML}}|\xi) \Big|_{\xi=\xi^{\text{ML}}+(\xi^{\text{ML}}-\xi)\Theta}. \end{aligned} \quad (89)$$

In analogy to Sect. 3.1 the Gaussian approximation is considered valid when the term of fourth order is negligible as compared to the term of second order for all $\xi - \xi^{\text{ML}}$ in the interval (38). For these values of ξ we have to find the maximum of the absolute value of the fourth derivative in Eq. (89). We call it $|H^{(4)}|_{\text{max}}$. In analogy to Eq. (42) the Gaussian approximation is accepted when

$$\frac{1}{2} \left(\frac{3\sigma}{\sqrt{N}} \right)^2 F \gg \frac{1}{24} \left(\frac{3\sigma}{\sqrt{N}} \right)^4 |H^{(4)}|_{\text{max}} \quad (90)$$

or

$$1 \gg \frac{3\sigma^2}{4N} \frac{|H^{(4)}|_{\text{max}}}{F}. \quad (91)$$

As in Eq. (39) the Fisher information F equals σ^{-2} and we obtain the condition

$$1 \gg \frac{3|H^{(4)}|_{\text{max}}}{4NF^2}. \quad (92)$$

In appendix F the value of $|H^{(4)}|_{\text{max}}$ is found to be

$$|H^{(4)}|_{\text{max}} = 16. \quad (93)$$

The value of the Fisher information F is given by Eq. (81). Whence, the condition (92) for the Gaussian approximation becomes

$$1 \gg \frac{3}{4N}. \quad (94)$$

We consider it to be fulfilled if

$$0.1 \geq \frac{3}{4N} \quad (95)$$

or

$$N \geq 8. \quad (96)$$

Thus the condition (96) for the Gaussian approximation to the binomial model is much more easily fulfilled than the corresponding condition (44) for the chi-squared model. The reason is: In Sect. 3.1 the remainder was of third order; in the present case it is of fourth order.

This is exemplified by FIG. 2:

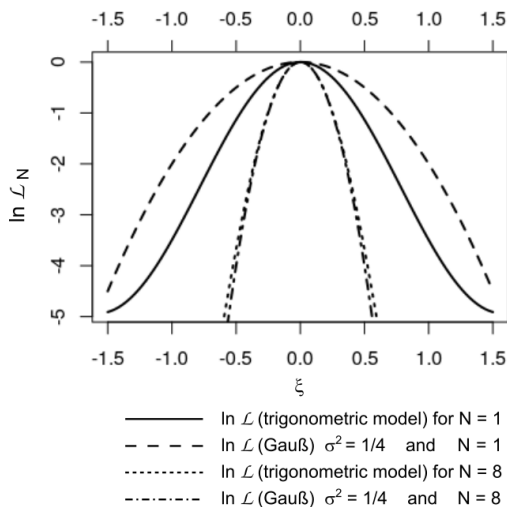


Figure 2. Comparison of the logarithmic likelihood function of the trigonometric model with form invariance to that of a Gaussian (i.e. a parabola) for $N = 1$ and $N = 8$.

In Fig. 2 we compare the logarithmic likelihood function of the trigonometric model with form invariance to that of a Gaussian (i.e. a parabola) for $N = 1$ and $N = 8$. The former is taken from Eq. (83), the latter from Eq. (27) with $\sigma^2 = F^{-1} = 1$. The curvatures of the two likelihood functions agree at their maxima. The figure also illustrates the reflection symmetry of the trigonometric model with respect to $\xi = \xi^{ML} = 0$. Note that with increasing N the likelihood functions are ever more concentrated around their maximum value.

The step from (94) to (95) is due to the common idea that a given positive x be large against $y > 0$ when x is one order of magnitude larger than y . This idea is commonly used for approximations in mathematics [8] and physics [10]. If a higher accuracy is required, this rule can easily be adapted.

In Chap. 12 of [25] a version of Item Response Theory is presented which makes use of the trigonometric IRF (60). In that context a (simulated) competence test was discussed which asked 20 questions. From the present result (96) follows that the estimated person-parameters have a Gaussian posterior distribution.

In Sect. 3.2 as well as in the present section the condition for the Gaussian approximation is independent of the value of ξ^{ML} . In the case of the binomial model, this means that the Gaussian approximation is valid even for uniform or close to uniform patterns of answers.

6 Conclusions

The present text presents a criterion for the validity of the Gaussian approximation to the likelihood function of a statistical model p when N observations $\mathbf{x} = (x_1, \dots, x_N)$ have been collected.

We require that a statistical model possesses a symmetry between the observed quantity x and the parameter ξ . This symmetry is defined in terms of a Lie group. It has been called form invariance. It allows to formally specify the prior distribution — required by Bayes but not specified by him. The

model p can then be parameterised such that it depends on the difference between the observed quantity x and the parameter ξ which conditions the observations. The model $p(x - \xi)$ remains invariant when both quantities are shifted by the same amount. We call this property translational form invariance. Then the likelihood function $\mathcal{L}_N(\xi)$ is shown to depend on the difference $\xi^{ML} - \xi$ between the maximum likelihood estimator ξ^{ML} and the parameter ξ . This means that the Bayesian posterior distribution, too, depends on $\xi^{ML} - \xi$. One can even shift the scale of ξ such that the ML estimator is found at $\xi^{ML} = 0$.

A total of N observations conditioned by one and the same parameter ξ is generally expected to lead to a Gaussian likelihood function for sufficiently large N . The question of how large N must be, in order to justify the Gaussian approximation, is answered in Sects. 3 and 5 for two quite different examples. The basic idea is, that a valid Gaussian approximation to the posterior distribution means that the error assigned to the parameter ξ equals $\sigma = (NF)^{-1/2}$ within the interval $|\xi^{ML} - \xi| < 3\sigma$. Here, F is the Fisher information yielded by the model p . The probability for ξ to lie outside this 3σ -interval is neglected. A stricter condition for the Gaussian approximation would be achieved by requiring an interval larger than 3σ for its validity.

The example of Sect. 3 is a version of the chi-squared model with two degrees of freedom. This distribution, as well as its posterior, has a considerable skewness. In this case one needs $N = 160$ observations for the Gaussian approximation to be acceptable. The large number of required observations is due to the skewness.

The example of Sect. 5 is the trigonometric model, a specific form of the binomial model based on form invariance. It strongly differs from the chi-squared model since the observations are not taken from a continuum of real numbers but rather from the alternative of 0 or 1. The likelihood function turns out to exhibit a mirror symmetry — as the Gaussian exhibits, too. This helps to approach the Gaussian distribution. We find: For $N \geq 8$ the posterior of the trigonometric model can be considered as Gaussian. This holds for every value of the ML estimator. It is a favorable result for the application of the binomial model to competence tests such as the PISA studies [46].

In general, we see a practical interest in our results since the normal distribution is the basis of parametric methods in applied statistics, widely used in many areas (education, medicine, science, etc.) [59, 60, 58]. To know whether the normal distribution is applicable or not, is of interest for practitioners in these fields.

A Comparing Two Distributions

We show that for any two normalised distributions p and q the inequality

$$\sum_l p_l \ln \frac{q_l}{p_l} \leq 0 \tag{97}$$

holds provided that p and q are labelled by the binary numbers l and are normalised according to

$$\begin{aligned}\sum_l q_l &= 1, \\ \sum_l p_l &= 1.\end{aligned}\quad (98)$$

The l.h.s. of (97) attains its maximum value of zero when and only when p_l and q_l agree for every l .

The inequality (97) is a consequence of the inequality

$$\ln s \leq s - 1 \quad \text{for } s > 0. \quad (99)$$

The linear function $s - 1$ is tangent to the function $\ln s$. The common point lies at $s = 1$, where both functions have the value of 1. Setting $s = q_l/p_l$ the inequality entails

$$\ln \frac{q_l}{p_l} \leq \frac{q_l}{p_l} - 1 \quad (100)$$

or

$$p_l \ln \frac{q_l}{p_l} \leq q_l - p_l \quad (101)$$

for every l . The quantity on the l.h.s. has also been introduced by Campbell in Sect. 1 of Chap. 5 of [9]. Summing (101) over l yields the inequality (97). When the distributions p and q agree with each other, the l.h.s. of (97) vanishes. Then and only then the expression assumes its maximum value.

One can interpret p_l as the probability $p(x_l - \xi^{\text{ML}}) \frac{\Delta x}{N}$ contained in a bin centered at the value x_l of the real variable x and having the width $\frac{\Delta x}{N}$. Here, p shall be a normalised probability density. Similarly one can interpret q_l as the probability $p(x_l - \xi) \frac{\Delta x}{N}$. In the limit of $\frac{\Delta x}{N} \rightarrow 0$ the inequality (97) then yields the inequality

$$\frac{1}{N} \int_{\text{domain } x} dx p(x - \xi^{\text{ML}}) \ln \frac{p(x - \xi)}{p(x - \xi^{\text{ML}})} \leq 0. \quad (102)$$

or

$$\begin{aligned}\int_{\text{domain } x} dx p(x - \xi^{\text{ML}}) \ln p(x - \xi) &\leq \\ \int_{\text{domain } x} p(x - \xi^{\text{ML}}) \ln p(x - \xi^{\text{ML}}).\end{aligned}\quad (103)$$

B The ML Estimator of the Chi-Squared Model

The ML estimator of the chi-squared model (45) is calculated.

Up to an additive constant (independent of ξ) the logarithmic likelihood function is given by

$$\ln \mathcal{L}_N(\xi) = \text{const} + [x - \xi - \exp(x - \xi)]. \quad (104)$$

The ML estimator ξ^{ML} solves the ML equation

$$\begin{aligned}0 &= \frac{d}{d\xi} \ln \mathcal{L}_N(\xi) \\ &= -1 + \exp(x - \xi)\end{aligned}\quad (105)$$

The solution is

$$\xi^{\text{ML}} = x. \quad (106)$$

C Two Versions of the Fisher Information

It shall be shown that the two lines of Eq. (30) agree with each other. Let us start from the second line which we write as

$$\frac{\partial^2}{\partial \xi^2} p(x|\xi) = \frac{\partial^2}{\partial \xi'^2} \int_{-\infty}^{\infty} dx p(x|\xi) \ln p(x|\xi') \Big|_{\xi'=\xi}. \quad (107)$$

This expression can be rewritten

$$\begin{aligned}F(\xi) &= - \frac{\partial}{\partial \xi'} \int dx p(x|\xi) \frac{\frac{\partial}{\partial \xi'} p(x|\xi')}{p(x|\xi')} \Big|_{\xi'=\xi} \\ &= - \int dx p(x|\xi) \left[\frac{\frac{\partial^2}{\partial \xi'^2} p(x|\xi')}{p(x|\xi')} - \left(\frac{\frac{\partial}{\partial \xi'} p(x|\xi')}{p(x|\xi')} \right)^2 \right]_{\xi'=\xi} \\ &= - \int dx \left[\frac{\partial^2}{\partial \xi^2} p(x|\xi) - p(x|\xi) \left(\frac{\partial}{\partial \xi} \ln p(x|\xi) \right)^2 \right] \\ &= - \frac{\partial^2}{\partial \xi^2} \int_{-\infty}^{\infty} dx p(x|\xi) + \int_{-\infty}^{\infty} dx p(x|\xi) \left(\frac{\partial}{\partial \xi} \ln p(x|\xi) \right)^2\end{aligned}\quad (108)$$

The first one of the two integrals in the last line vanishes since p is normalised to unity for every ξ . The second integral in the last line corresponds to the first line of Eq. (30).

D The Chi-Squared Model

Each of the quantities x_k , where $k = 1, 2$, shall have the Gaussian distribution

$$w(x_k|\sigma) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{x_k^2}{2\sigma^2}\right), \quad -\infty < x_k < \infty, \quad (109)$$

with one and the same root mean square value σ . The chi-squared model $\chi_2^{\text{sq}}(T|\sigma)$ with two degrees of freedom is the distribution of the quantity

$$T = x_1^2 + x_2^2. \quad (110)$$

It is given by

$$\chi_2^{\text{sq}}(T|\sigma) = \frac{1}{2\sigma^2\Gamma(1)} \exp\left(-\frac{T}{2\sigma^2}\right), \quad 0 < T, \tau < \infty, \quad (111)$$

see e.g. Eq. (4.34) of Ref. [25]. This distribution is normalised to unity. The transformations

$$\begin{aligned}z &= \ln T, \\ \zeta &= \ln(2\sigma^2)\end{aligned}\quad (112)$$

lead to

$$\begin{aligned}\tilde{\chi}_2^{\text{sq}}(z|\zeta) &= \frac{dT}{dz} \chi_2^{\text{sq}}(T|\sigma) \\ &= T \chi_2^{\text{sq}}(T|\sigma),\end{aligned}\quad (113)$$

where T and σ must be expressed by z and ζ . This gives

$$\tilde{\chi}_2^{\text{sq}}(z|\zeta) = \exp\left(z - \zeta - e^{z-\zeta}\right) \quad (114)$$

which corresponds to Eq. (45).

E Derivatives of the Functional H for the chi-Squared Distribution

The integral in the first line of Eq. (50) yields $\Gamma(2)$. It is a special case of the formula

$$\Gamma(z) = \int_{-\infty}^{\infty} dt \exp(zt - e^t) \tag{115}$$

given in Sect. 8.312, no. 10 of Ref. [22]. The value of the Gamma function required in Eq. (50) is

$$\Gamma(2) = 2. \tag{116}$$

F The Functional H for the Trigonometric Model with Translational Invariance

The functional $H(\xi^{ML}|\xi)$ for the trigonometric model model (75) is given by Eq. (83) to be

$$H(\xi^{ML}|\xi) = \mathcal{N} \int_{-\pi/2}^{\pi/2} ds \cos^2(s - \xi^{ML}) \ln \frac{\cos^2(s - \xi)}{\cos^2(s - \xi^{ML})}. \tag{117}$$

The substitution

$$s' = s - \xi^{ML} \tag{118}$$

yields

$$\begin{aligned} & H(\xi^{ML}|\xi) \\ &= \mathcal{N} \int_{-\pi/2+\xi^{ML}}^{\pi/2+\xi^{ML}} ds' \cos^2 s' \\ & \quad \ln \frac{\cos^2(s' - \xi + \xi^{ML})}{\cos^2 s'} \\ &= \mathcal{N} \int_{-\pi/2}^{\pi/2} ds' \cos^2 s' \\ & \quad \ln \frac{\cos^2(s' - \xi + \xi^{ML})}{\cos^2 s'} \\ &= H(0|\xi - \xi^{ML}). \end{aligned} \tag{119}$$

Here, the second line is obtained from the first one because the integrand is periodic with a period of π , hence, the shift of the limits of integration is immaterial. With the abbreviation

$$\alpha = \xi^{ML} - \xi \tag{120}$$

this reads

$$\begin{aligned} H(\xi^{ML}|\xi) &= H(0|-\alpha) \\ &= \mathcal{N} \int_{-\pi/2}^{\pi/2} ds \cos^2 s \ln \frac{\cos^2(s + \alpha)}{\cos^2 s} \end{aligned} \tag{121}$$

Substituting

$$s' = -s \tag{122}$$

in the integral (121) one obtains

$$\begin{aligned} H(0|-\alpha) &= -\mathcal{N} \int_{\pi/2}^{-\pi/2} ds' \cos^2 s' \ln \frac{\cos^2(-s' + \alpha)}{\cos^2 s'} \\ &= \mathcal{N} \int_{-\pi/2}^{\pi/2} ds' \cos^2 s' \ln \frac{\cos^2(s' - \alpha)}{\cos^2 s'} \\ &= H(0|\alpha). \end{aligned} \tag{123}$$

Although $\cos(s - \alpha)$ vanishes at a point within the domain of integration, the integral (123) exists and can be obtained as if the integrand were simply undefined at this isolated point, see appendix H.

Comparing (121) with (123) shows that $H(0|-\alpha)$ is a mirror-symmetrical function of the difference α . By consequence, all odd derivatives vanish at $\alpha = 0$,

$$\left. \frac{\partial^{2\nu+1}}{\partial \alpha^{2\nu+1}} H(0|-\alpha) \right|_{\alpha=0} = 0, \quad \nu = 0, 1, 2, \dots \tag{124}$$

We calculate the derivatives with respect to α . The first derivative is the basis of all higher ones. It must be rewritten in order to see that all derivatives exist. Starting from Eq. (123) we find

$$\begin{aligned} & \frac{\partial}{\partial \alpha} H(0|-\alpha) \\ &= \mathcal{N} \frac{\partial}{\partial \alpha} \int_{-\pi/2}^{\pi/2} ds [\cos^2 s \ln \cos^2(s - \alpha) - \cos^2 s \ln \cos^2 s] \\ &= \mathcal{N} \frac{\partial}{\partial \alpha} \int_{-\pi/2}^{\pi/2} ds \cos^2 s \ln \cos^2(s - \alpha) \end{aligned} \tag{125}$$

By use of the substitution

$$s' = s - \alpha \tag{126}$$

one obtains from the integral (125)

$$\begin{aligned} & \frac{\partial}{\partial \alpha} H(0|-\alpha) \\ &= \mathcal{N} \frac{\partial}{\partial \alpha} \int_{-\pi/2}^{\pi/2} ds \cos^2(s + \alpha) \ln \cos^2 s \\ &= 2\mathcal{N} \frac{\partial}{\partial \alpha} \int_{-\pi/2}^{\pi/2} ds \cos^2(s + \alpha) \ln \cos s. \end{aligned} \tag{127}$$

We express $\cos(s + \alpha)$ by the sum

$$\cos(s + \alpha) = \cos s \cos \alpha - \sin s \sin \alpha \tag{128}$$

and obtain

$$\begin{aligned} & \frac{\partial}{\partial \alpha} H(0|-\alpha) \\ &= 2\mathcal{N} \frac{\partial}{\partial \alpha} \int_{-\pi/2}^{\pi/2} ds [\cos s \cos \alpha - \sin s \sin \alpha]^2 \ln \cos s. \end{aligned} \tag{129}$$

The square $[\dots]^2$ of a binomial expression displays two squares and a mixed term. Here, the mixed term, as a function of s , is antisymmetric with respect to $s = 0$. Therefore the integral over the mixed term vanishes and we obtain

$$\begin{aligned} & \frac{\partial}{\partial \alpha} H(0|-\alpha) \\ &= 2\mathcal{N} \frac{\partial}{\partial \alpha} \int_{-\pi/2}^{\pi/2} ds [\cos^2 s \cos^2 \alpha + \sin^2 s \sin^2 \alpha] \ln \cos s. \end{aligned} \tag{130}$$

This integrand, as a function of s , is symmetric with respect to $s = 0$. Therefore we have

$$\begin{aligned} & \frac{\partial}{\partial \alpha} H(0 | -\alpha) \\ &= 4\mathcal{N} \frac{\partial}{\partial \alpha} \cos^2 \alpha \int_0^{\pi/2} ds \cos^2 s \ln \cos s \\ & \quad + 4\mathcal{N} \frac{\partial}{\partial \alpha} \sin^2 \alpha \int_0^{\pi/2} ds \sin^2 s \ln \cos s \\ &= 8\mathcal{N} \cos \alpha \sin \alpha \left[- \int_0^{\pi/2} ds \cos^2 s \ln \cos s \right. \\ & \quad \left. + \int_0^{\pi/2} ds \sin^2 s \ln \cos s \right]. \end{aligned} \quad (131)$$

Let us introduce the integrals

$$I_0 = \int_0^{\pi/2} ds \ln \cos s \quad (132)$$

and

$$I_2 = \int_0^{\pi/2} ds \sin^2 s \ln \cos s. \quad (133)$$

They allow to write Eq. (131) as

$$\frac{\partial}{\partial \alpha} H(0 | -\alpha) = 8\mathcal{N} [-I_0 + 2I_2] \cos \alpha \sin \alpha \quad (134)$$

which gives

$$\frac{\partial}{\partial \alpha} H(0 | -\alpha) = -4\mathcal{N} [I_0 - 2I_2] \sin(2\alpha) \quad (135)$$

by help of the identity

$$\sin(2\alpha) = 2 \cos \alpha \sin \alpha. \quad (136)$$

The derivative (135) vanishes at $\alpha = 0$, as expected from Eq. (124).

The values of the integrals (132) and (133) can be taken from the table of integrals [22]. According to the entries 4.387 no. 3 and 8.366 no. 1.2 of [22], the integral (132) has the value

$$I_0 = -\frac{\pi}{2} \ln 2 \quad (137)$$

while the entries 4.387 no. 8 and 8.365 no. 1 as well as 8.366 no.1 yield

$$I_2 = -\frac{\pi}{8} [2 \ln 2 + 1]. \quad (138)$$

These two values lead to

$$I_0 - 2I_2 = \frac{\pi}{4}. \quad (139)$$

Together with the value of \mathcal{N} in Eq. (77) the derivative (135) becomes

$$\frac{\partial}{\partial \alpha} H(0 | -\alpha) = -2 \sin(2\alpha). \quad (140)$$

Therefore the second derivative of $H(0 | -\alpha)$ becomes

$$\frac{\partial^2}{\partial \alpha^2} H(0 | -\alpha) = -4 \cos(2\alpha) \quad (141)$$

and the fourth derivative becomes

$$\frac{\partial^4}{\partial \alpha^4} H(0 | -\alpha) = 16 \cos(2\alpha). \quad (142)$$

The maximum of the absolute value of the fourth derivative is found at $\alpha = 0$ which means

$$|H^{(4)}|_{\max} = 16. \quad (143)$$

Note that the second derivative in Eq. (141) for $\alpha = 0$ gives the value

$$\frac{\partial^2}{\partial \alpha^2} H(0 | \alpha) \Big|_{\alpha=0} = -4 \quad (144)$$

which is, up to its sign, the Fisher information (81) of the trigonometric model with translational invariance.

G The Normalisation of the Trigonometric Model with Translational Invariance

By partial integration of the \cos^2 -function one finds

$$\int_{-\pi/2}^{\pi/2} ds \cos^2 s = [\sin s \cos s]_{-\pi/2}^{\pi/2} + \int_{-\pi/2}^{\pi/2} ds \sin^2 s. \quad (145)$$

This leads to

$$\int_{-\pi/2}^{\pi/2} ds \cos^2 s = \int_{-\pi/2}^{\pi/2} ds (1 - \cos^2 s) \quad (146)$$

or

$$2 \int_{-\pi/2}^{\pi/2} ds \cos^2 s = \pi \quad (147)$$

which proves Eq. (77).

H Integrating over a Logarithmic Divergence

In the ϵ -interval

$$-\epsilon < s' - s'_0 < \epsilon, \quad \epsilon > 0, \quad (148)$$

around the point s'_0 , where \cos^2 vanishes, the \cos^2 -function behaves as

$$\cos^2 s' = c(s' - s'_0)^2 \quad (149)$$

since the \cos^2 -function does not become negative and ϵ can be chosen arbitrarily small. Here, c is a positive number. We show that the integral over the ϵ -interval

$$\int_{s'_0-\epsilon}^{s'_0+\epsilon} ds' \ln \cos^2 s' = 2\epsilon \ln c + \int_{s'_0-\epsilon}^{s'_0+\epsilon} ds' \ln(s' - s'_0)^2 \quad (150)$$

exists.

For this we rewrite

$$\begin{aligned} & \int_{s'_0-\epsilon}^{s'_0+\epsilon} ds' \ln(s' - s'_0)^2 \\ &= \int_{s'_0-\epsilon}^{s'_0} ds' \ln(s' - s'_0)^2 + \int_{s'_0}^{s'_0+\epsilon} ds' \ln(s' - s'_0)^2 \\ &= 2 \int_{s'_0-\epsilon}^{s'_0} ds' \ln(s'_0 - s') + 2 \int_{s'_0}^{s'_0+\epsilon} ds' \ln(s' - s'_0) \end{aligned} \quad (151)$$

so that the arguments of the logarithms are non-negative. In the first integral of the second line we substitute

$$s'_0 - s' = x \tag{152}$$

and obtain

$$\begin{aligned} 2 \int_{s'_0-\epsilon}^{s'_0} ds' \ln(s'_0 - s') &= -2 \int_{\epsilon}^0 dx \ln x \\ &= 2 \int_0^{\epsilon} dx \ln x \\ &= 2 [x \ln x - 1]_{x=0}^{x=\epsilon} \\ &= 2\epsilon \ln \epsilon. \end{aligned} \tag{153}$$

In a similar way one finds the same result for the second integral on the second line of (151). Thus Eq. (150) yields

$$\int_{s'_0-\epsilon}^{s'_0+\epsilon} ds' \ln \cos^2 s'_0 = 2\epsilon \ln c + 4\epsilon \ln \epsilon, \tag{154}$$

and this contributes a negligible amount to the expression (82) when ϵ is small. Thus $\ln \mathcal{L}_N$ and the functional $H(\xi^{\text{ML}}|\xi)$ exist.

I The Likelihood Function of a Gaussian Model

The N -fold Gaussian model

$$G_N(\mathbf{x}|\xi) = (2\pi\sigma^2)^{-N/2} \prod_{k=1}^N \exp\left(-\frac{(x_k - \xi)^2}{2\sigma^2}\right) \tag{155}$$

is given in Eq. (7). We write it as

$$\begin{aligned} &G_N(\mathbf{x}|\xi) \\ &= (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^N (x_k^2 - 2x_k\xi + \xi^2)\right) \\ &= (2\pi\sigma^2)^{-N/2} \\ &\exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^N (\langle x^2 \rangle - 2\langle x \rangle \xi + \xi^2)\right) \end{aligned} \tag{156}$$

by introducing the averages

$$\begin{aligned} \langle x^2 \rangle &= \frac{1}{N} \sum_{k=1}^N x_k^2, \\ \langle x \rangle &= \frac{1}{N} \sum_{k=1}^N x_k. \end{aligned} \tag{157}$$

It is not difficult to factorise this according to

$$G_N(\mathbf{x}|\xi) = (2\pi\sigma^2)^{-N/2} \left(\exp\left(-\frac{N}{2\sigma^2} (\langle x^2 \rangle - \langle x \rangle^2)\right) \exp\left(-\frac{N}{2\sigma^2} (\langle x \rangle - \xi)^2\right) \right).$$

The posterior distribution $G_N(\xi|\mathbf{x})$ is given by the factor that depends on ξ , i.e.

$$G_N(\xi|\mathbf{x}) \propto \exp\left(-\frac{N}{2\sigma^2} (\langle x \rangle - \xi)^2\right) \tag{158}$$

in agreement with Eq. (8). The maximum of the likelihood function occurs at

$$\xi^{\text{ML}} = \langle x \rangle. \tag{159}$$

J The Fisher Information of the Binomial Model

The Fisher information of the model binomial model (59) shall be independent of ξ . This means that the expression

$$\begin{aligned} F(\xi) &= \sum_{x=0}^1 q(x|\xi) \left[\frac{\partial}{\partial \xi} \ln q(x|\xi) \right]^2 \\ &= [1 - R] \left[\frac{R'}{1 - R} \right]^2 + R \left[\frac{R'}{R} \right]^2 \\ &= \frac{[R']^2}{1 - R} + \frac{[R']^2}{R} \\ &= \frac{[R']^2}{R[1 - R]} \end{aligned} \tag{160}$$

be independent of ξ . Here, R' is the derivative of $R(\xi)$. Thus the numerator of (160) should be proportional to the denominator. This is reached when we set

$$R(\xi) = \cos^2 \xi, \quad -\pi/2 \leq \xi \leq \pi/2 \tag{161}$$

and obtain

$$F(\xi) \equiv 4. \tag{162}$$

REFERENCES

- [1] J. Frank Adams. *Lectures on Lie Groups*. W.A. Benjamin, New York, 1969.
- [2] John Aldrich. R.A. Fisher and the Making of the Maximum Likelihood 1912–1922. *Statistical Science*, 12(3):172–176, 1997.
- [3] A. Aykac and C. Brumat, editors. *New Methods in the Applications of Bayesian Methods*, volume 1. North Holland, Amsterdam, 1977.
- [4] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, 53:330–418, 1763. Reprinted in *Biometrika* 45, 293–315 (1958) and in ‘Studies in the History of Statistics and Probability’ E.S. Pearson and M.G. Kendall eds., C.Griffin and Co. Ltd., London 1970 and in ‘Two papers by Bayes with commentaries’ W.E. Deming ed., Hafner Publishing Co., New York, 1963.
- [5] J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M Smith, editors. *Bayesian Statistics 3. Proceedings of the third International Meeting held at Valencia on June 1–5, 1987*, Oxford, 1988. Oxford University Press.

- [6] Andrew C. Berry. The accuracy of the Gaussian approximation to the sum of independent variables. *Transactions of the American Mathematical Society*, 49:122–136, 1941.
- [7] Theodor Bröcker and Tammo tom Dieck. *Representations of Compact Lie Groups*. Springer, New York, 1985.
- [8] I.N. Bronstein, K.A. Semendyayev, G. Musiol, and H. Muehlig. *Handbook of Mathematics*. Springer, New York, 4th edition, 2014.
- [9] L. Lore Campbell. Geometric ideas in minimum cross-entropy. In Jawaharlal Karmeshu, editor, *Entropy measures, maximum entropy principle, and emerging applications*, chapter 5, pages 103–114. Springer, Berlin, 2003.
- [10] K. Cummings, P.W. Laws, E.F. Redish, and P.J. Cooney. *Understanding Physics*. John Wiley and Sons, Hoboken N.J., 2004.
- [11] A.W.F. Edwards. *Likelihood*. Johns Hopkins University Press, Baltimore, 1992.
- [12] George A. Ferguson. Item selection by the constant process. *Psychometrika*, 7(1):19–29, 1942.
- [13] G.H. Fischer. On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, 46(1):59–77, 1981.
- [14] G.H. Fischer. Derivations of the Rasch model. In *Rasch models: foundations, recent developments and applications*, pages 15–38, New York, 1995. Workshop held at the University of Vienna, Feb. 25–27, 1993.
- [15] R.A. Fisher. An absolute criterion for fitting frequency curves. *Messenger of Math.*, 41:155–160, 1912.
- [16] R.A. Fisher. On the mathematical foundations of theoretical statistics. *Philos. Trans. Royal Soc. London Ser. A*, 222:309–368, 1922.
- [17] R.A. Fisher. Theory of statistical information. *Proc. Cambridge Philos. Soc.*, 22:700–725, 1925.
- [18] R.A. Fisher. *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh, 2nd edition, 1959.
- [19] R.A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1967. 13-th edition.
- [20] P.F. Fougère, editor. *Maximum Entropy and Bayesian Methods, Dartmouth, 1989*, Dordrecht, 1990. Kluwer.
- [21] Christoph Fuhrmann. *Eine trigonometrische Parametrisierung von Kompetenzen. Zur Methodologie der probabilistischen Bildungsforschung*. PhD thesis, Ruhr-Universität, Bochum, Germany, 2016. This dissertation is published by Springer VS, Wiesbaden (Germany), 2017.
- [22] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, New York, eighth edition, 2015. Edited by D. Zwillinger.
- [23] Louis Guttman. *The basis of scalogram analysis*, pages 60–90. Smith, Gloucester, Mass., 1973.
- [24] A. Haar. Der Maßbegriff in der Theorie der kontinuierlichen Gruppen. *Ann. Math.*, 34:147–169, 1933.
- [25] H.L. Harney. *Bayesian Inference. Parameter Estimation and Decisions*. Springer, Heidelberg, 2016. Second edition.
- [26] J. Hartigan. Invariant prior distributions. *Ann. Math. Statist.*, 35:836–845, 1964.
- [27] A. Hobson and B.K. Cheng. A comparison of the Shannon and Kullback information measures. *J. of Statistical Physics*, 7(4):301–310, 1973.
- [28] R.V. Hogg, E.A. Tanis, and D.L. Zimmermann. *Probability and Statistical Inference*. Pearson, Boston, 2015.
- [29] Shun ichi Amari. *Differential Geometrical Methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer, Heidelberg, 1985.
- [30] E.T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4(3):227–241, September 1968.
- [31] E.T. Jaynes and G.L. Bretthorst. *Probability theory: The logic of Science*. Cambridge Univ. Press, Cambridge, Mass., 2012.
- [32] H. Jeffreys. An invariant form of the prior probability in estimation problems. *Proc. of the Roy. Soc.*, 186 A:453–461, 1946.
- [33] R.E. Kass. *The Riemannian Structure of Model Spaces: A Geometrical Approach to Inference*. PhD thesis, University of Chicago, 1980. See especially pages 94–95.
- [34] R.E. Kass. Data translated likelihood and Jeffreys’s rules. *Biometrika*, 77:107–114, 1990.
- [35] R.E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *J. Am. Statist. Assoc.*, 91:1343–1370, 1996.
- [36] Maurice Kendall and Alan Stuart. *The Advanced Theory of Statistics*, volume 2. Griffin, London, third edition, 1951. For the notion of sufficient statistic see p. 22.
- [37] Maurice Kendall and Alan Stuart. *The Advanced Theory of Statistics*. Griffin, London, fourth edition, 1976.
- [38] S. Kullback and R.A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22:79–86, 1951.
- [39] Pierre Simon de Laplace. Mémoire sur la probabilité des causes par les événements. *Mém. de math. et phys. présentés à l’Acad. roy. des sci.*, 6:621–656, 1774. Reprinted in [40], vol. 8, pages 27–65. An English translation can be found in [57].
- [40] Pierre Simon de Laplace. *Œuvres complètes de Laplace*. Gauthier-Villars, Paris, 1886–1912. 14 volumes.
- [41] Paul F. Lazarsfeld. The logical and mathematical foundation of latent structure analysis. In Samuel A. Stouffer and Louis Guttman et al., editors, *Measurement and Prediction*. Smith, Gloucester, Mass., reprinted edition, 1973.
- [42] E.L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New York, 1998.
- [43] Thurstone L.L. A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16(October):433–451, 1925.
- [44] Frederic M. Lord. *Applications of item response theory to practical testing problems*. Erlbaum, Hillsdale NJ, 1980.

- [45] J. Neyman and Elizabeth L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32, 1948.
- [46] OECD. PISA 2015. Technical report. www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf. This website was visited on November 3, 2018.
- [47] A. Papoulis and S.U. Pillai. *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, 2002.
- [48] C. Radakrishna Rao. Differential metrics in probability spaces based on entropy and divergence measures. Technical Report Report No. AD-A160301, AFOSR-TR-85-0864, Air Force Office of Scientific Research, Bolling AFB, DC, 1985.
- [49] G. Rasch. An individual approach to item analysis. In P.F. Lazarsfeld and N.W. Henry, editors, *Readings in mathematical social science*, pages 89–108, Cambridge, 1966. Massachusetts Institute of Technology.
- [50] G. Rasch. An informal report on a theory of objectivity in comparisons. In L.J.Th. van der Kamp and C.A.J. Vlek, editors, *Measurement theory. Proceedings of the NUFFIC international summer session in science at 'Het Oude Hof'. The Hague, July 14–28, 1966*, pages 1–19, Leiden, 1967. University of Leiden.
- [51] G. Rasch. On specific objectivity: An attempt at formalising the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, 14:58–94, 1977.
- [52] G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. The University of Chicago Press, Chicago, 1980.
- [53] Georg Rasch. *Probabilistic models for some intelligence and attainment tests with a foreword and afterword by Benjamin D. Wright*. Danish Institute for Educational Research, 1960. expanded edition 1980 published by University of Chicago Press.
- [54] S.M. Ross. *Introductory Statistics*. Elsevier/AP, Amsterdam, 2017.
- [55] Jürgen Rost. *Lehrbuch Testtheorie — Testkonstruktion*. Hans Huber, Bern, second revised edition, 2004.
- [56] Aris Spanos. Revisiting the Neyman-Scott model: an inconsistent MLE or an ill-defined model?, 2013. arXiv:1301.6278v1 [stat.ME] 26 January 2013.
- [57] S.M. Stigler. Laplace's 1774 memoir on inverse probability. *Statist. Sci.*, 1:359–378, 1986.
- [58] Arif Rahman, Oke Oktavianty, Ratih Ardia Sari, Wifqi Azlia and Lavestya Dina Anggreni. Gaussian Distribution on Validity Testing to Analyze the Acceptance Tolerance and Significance Level. *Mathematics and Statistics*, Vol. 8, No. 1, pp. 42 - 52, 2020. DOI: 10.13189/ms.2020.080105.
- [59] Johnson, N. L., Kotz, S. and Balakrishnan, N. *Continuous univariate distributions*, ch. 13.2. New York [N.Y., etc.: Wiley]. 1994
- [60] King, A. P. and Eckersley, R. J. *Statistics for biomedical engineers and scientists: How to visualize and analyze data*. London: Academic Press, ch. 4.5. 2019
- [61] S.M. Stigler. The epic story of maximum likelihood. *Statistical Science*, 22(4):598–620, 2007.
- [62] Arnold Zellner. *Maximal Data Information Prior Distributions*, chapter 12. Volume 1 of Aykac and Brumat [3], 1977.
- [63] A. Zellner. *An introduction to Bayesian inference in econometrics*. Wiley, New York, 1996.