Journal of Convex Analysis
Volume 29 (2022), No. 1, 1–xx

Noncentral distributions, entropic convex dual, Gibbs priors & orthogonal polynomials

Richard Le Blanc*
FMSS, Université de Sherbrooke
3001, 12 ème Avenue Nord
Sherbrooke, Québec, Canada J1H 5N4
richard.le.blanc@usherbrooke.ca

Received: January 1, 2022
Accepted: December 31, 2022

In parametric Bayesian inference, data-constrained Jaynes maximal entropy priors can be computed objectively whenever dense datasets are available. The maximal entropy is obtained by minimizing the Gibbs potential, yielding Gibbs priors built upon the entropic convex dual $\lambda(r)$ of the empirical density $\rho(r)$ to be modelled. The four noncentral t, normal, F, and χ^2 univariate distributions $\rho(r|r_o)$, with r_o the noncentrality parameter of the respective distributions, can all be constructed in a modular fashion by multiplying their central counterparts $\rho(r|r_o = 0)$ with a factor $T(r|r_o)$ effecting a central distribution translation, that is, $\rho(r|r_o) = T(r|r_o)\rho(r|0)$. These multiplicative translation factors T have interesting analytical properties and useful applications. The translation factors $T(r|r_o)$ for the noncentral ultraspherical t distribution and the noncentral normal distribution are generating functions for the Gegenbauer and Hermite orthogonal polynomials, respectively, while the translation factors for the noncentral ultraspherical F distribution and the noncentral χ^2 distribution can be expanded in terms of the Jacobi and Laguerre orthogonal polynomials, respectively. This allows for expansion of the entropic convex dual $\lambda(r)$ on a small number of low order orthogonal polynomials, which greatly curtails the computational cost of its determination. When the factor $T(r|r_o)$ is weighted with an appropriate prior $\pi(r_o)$ and reexpressed in terms of the central distribution p-value, the resulting Bayes factor $BF(p)$ models its argument density $\rho(p) = BF(p)$ for the corresponding generic density $\rho(r) = \int \rho(r|r_o)\pi(r_o)\mathrm{d}r_o$, and allows computation of a local false discovery rate $\mathrm{fdr}(p) = 1/(1 + BF(p))$. Analysis of a genome-wide association study (GWAS) dataset illustrates the strengths of this orthogonal polynomial-based approach.

Keywords: Noncentral distributions, Orthogonal polynomials, Bayesian inference, Jaynes maximal entropy principle, Gibbs prior, Entropic convex dual.

2010 Mathematics Subject Classification: 05E35, 62F15, 94A17, 46N10.

1. Introduction

The four noncentral univariate t, F, normal, and χ^2 distributions $\rho(r|r_o)$, with r_o the noncentrality parameter of the respective distributions, can all be constructed in a modular fashion by multiplying their central counterparts $\rho(r|r_o = 0)$ with a factor $T(r|r_o)$ effecting a central distribution translation, that is,

$$\rho(r|r_o) = T(r|r_o)\ \rho(r|0). \tag{1}$$

* Corresponding author.

ISSN 0944-6532 / $ 2.50

With the exception of the normal distribution, these translations are non-shape preserving. This modular construction has direct applications in parametric Bayesian inference. Since the translation factor $T(r|r_o)$ stands for the likelihood ratio between the respective noncentral and central distributions, it can be weighted by a Bayesian prior $\pi(r_o)$ to obtain the Bayes factor

$$\mathrm{BF}(r) = \int T(r|r_o)\ \pi(r_o)\ \mathrm{d}r_o \tag{2}$$

for the generic superposition density

$$\rho(r) = \int \rho(r|r_o)\ \pi(r_o)\ \mathrm{d}r_o = \int T(r|r_o)\ \rho(r|0)\ \pi(r_o)\ \mathrm{d}r_o = BF(r)\ \rho(r|0). \tag{3}$$

According to the probability integral transform, the $p-$value distribution for the central distribution $\rho(r|0)$ is the uniform density $U(0,1)$ on the range $p = [0,1]$. Under the transformation $p = p(r)$, one therefore has that

$$\rho(p) = U(0,1) \times BF(p) = BF(p), \tag{4}$$

that is, the Bayes factor $BF(p)$ stands for the generally nonuniform $p-$value distribution $\rho(p)$ of the above generic superposition density. Whenever proper inferential priors can be provided, knowledge of the modular translation factors $T(r|r_o)$ thus allow one to go beyond the Null Hypothesis Statistical Testing (NHST) framework which only considers the central distribution with its uninformative uniform $p-$value distribution. Derivation of useful expressions for the modular noncentral distribution translation factors T then becomes a relevant undertaking.

Data-constrained Jaynes maximal entropy Bayesian priors can be computed objectively whenever dense datasets are available. The maximal entropy is reached by minimizing the Gibbs potential. The solution to this optimization problem requires determination of the entropic convex dual $\lambda(r)$ of the empirical density $\rho(r)$. See Le Blanc (2022) for an extensive review. Without introduction of proper analytical tools, one needs to determine $\lambda(r)$ on the full support domain of $\rho(r)$, a task which can be computationally expensive. Unfortunately, the classical approach to derivation of the noncentral univariate distributions does not provide us with the needed analytical tools. Recall that the noncentral t distribution has been classically derived by considering the ratio of a random variable distributing according to a normal distribution $N(\delta, 1)$ with non-vanishing mean δ over that of a random variable distributing according to a central χ_{ν_2} distribution with ν_2 degrees of freedom. See Gorroochurn (2016) for an historical perspective on this classical approach. Similarly, the noncentral F distribution is classically derived by considering the ratio of a noncentral χ^2 random variable with noncentrality parameter Λ and ν_1 degrees of freedom over that of a central χ^2 random variable with ν_2 degrees of freedom. The classical approach places primacy on the normal and χ^2 distributions, and the derivations produce complicated expressions for the noncentral distributions without any obvious geometrical interpretations for their constitutive components nor generalization properties (Van Aubel & Gawronski, 2003). Furthermore, these classical noncentral distributions all have a submodular decomposition for their translation factor of the form $T(r|r_o) = E(r|r_o)\ e^{-r_0^2/2}$, with $E(r|r_o)$ a generalized exponential, which does not provide for easy regrouping of terms of similar order in the noncentrality parameter r_o. Such a regrouping would allow expansion of the entropic dual convex $\lambda(r)$ on a small number of low order orthogonal polynomials (Szegö, 1939; Ismail *et al.*, 2005) which could greatly curtail the computational cost of determining the entropic convex dual $\lambda(r)$.

We chose to place primacy on the simple uniform density on high-dimensional hyperspheres $\mathcal{S}^n$ rather than on the normal and χ^2 distributions to derive surrogate noncentral distributions. It is known that the projection of a unit radius hypersphere $\mathcal{S}^n$ uniform density on any given axis — projection which readily provides us with the central t distribution — converges to that of a central normal distribution $N(0, 1)$ with null mean when the degree of freedom n tends to infinity (Vershynin, 2018, p. 59). This observation provides us with the needed building principle: use the high-dimensional hypersphere geometrical properties to derive modular expressions for the central and noncentral t, F, N, and χ^2 distributions providing both a geometrical interpretation for their modular components, and a proper framework to introduce relevant families of orthogonal polynomials. In order to distinguish the surrogate hypersphere-derived t and F distributions from the ones derived classically from the normal and χ^2 distributions, we convene hereafter to designate the former densities by the Greek letter υ (upsilon) as in υπερσφαίρα (ypersfaíra), that is, hypersphere in English, and the latter densities by the Greek letter ρ. It can be argued that the hyperspherical and the classical noncentral t and F distributions differ only through their translation factors T. A further point of nomenclature needs to be clarified: in the theory of orthogonal polynomials, the designation ultraspherical polynomials (also known as Gegenbauer polynomials (Olver *et al.*, 2021)) has prevailed over that of hyperspherical polynomials, even though the particle *ultra* is still being translated as υπερ (yper) in Greek. We shall therefore abide to this nomenclature in the following.

We will demonstrate that the translation factors $T(r|r_o)$ for the noncentral ultraspherical t distribution and the noncentral normal distribution are generating functions of the Gegenbauer and Hermite orthogonal polynomial families respectively, while the translation factors for the noncentral ultraspherical F distribution and the noncentral χ^2 distributions can be expanded in terms of the Jacobi and Laguerre orthogonal polynomial families respectively. This will allow expansion of the entropic convex dual $\lambda(r)$ on a small number of low order orthogonal polynomials, greatly curtailing the computational cost of its determination. In all cases, the respective central distribution $\rho(r|0)$ plays the role of defining weight $w(r)$ for the associated orthogonal polynomial family (Ismail *et al.*, 2005; Szegö, 1939).

In order to proceed, one needs to master some simple notions concerning the hypersphere geometry. The projection of a random unit vector $\mathbf{x}$ on the unit hypersphere $\mathcal{S}^{\nu_2}$ on any chosen unit polar axis $\mathbf{p}$ naturally defines a polar angle θ through the scalar product $\cos\theta = \mathbf{x} \cdot \mathbf{p}$. As such, the central t distribution with ν_2 degrees of freedom can be drawn on the compact support $-1 \leq \cos\theta \leq 1$, on which it acquires a simple expression in trigonometric terms: it is simply proportional to $\sin^{\nu_2-1}\theta$. On the familiar sphere $\mathcal{S}^2$ in 3D, the latter provides us with the well-known spherical surface element $\sin\theta\ \mathrm{d}\theta$ after integration of the azimuthal coordinate. Similarly, the central $F_{(\nu_1,\ \nu_2)}$ distribution becomes proportional to $\cos^{\nu_1-1}\theta\ \sin^{\nu_2-1}\theta$ on the compact domain $0 \leq \theta \leq \pi/2$, where θ is the angle between a random vector $\mathbf{x}$ on the hypersphere $\mathcal{S}^{\nu_1+\nu_2}$ and a secant hyperplane defining the subspace $\mathcal{S}^{\nu_1}$. The $\mathcal{S}^{\nu_1}$ and $\mathcal{S}^{\nu_2}$ subspaces will refer herein to the between-class and within-class variance spaces in an ANOVA, with $\nu_1 = c-1$, c being the number of classes, and $\nu_2 = n-c$, with n being the total number of samples. These two central ultraspherical distributions are all that is needed to proceed with derivation of the noncentral ultraspherical t, F, N, and χ^2 distributions where, for uniformity of designation, a normal distribution $N(\delta, 1)$ with non-vanishing mean δ and unit variance will be simply referred to as the noncentral normal distribution.

The manuscript has a simple and repetitive structure, except for three illustrative cases of increasing complexity. The translation factor $T(r|r_o)$ for the noncentral ultraspherical t distribution will be shown in Section 2 to be a generating function for modified Gegenbauer orthogonal polynomials, with corresponding weight function provided by the central ultraspherical t distribution. Determination of the entropic convex dual λ in terms of an expansion in a small number of low order Gegenbauer polynomials will be carried through in Section 3. The translation factor for the noncentral normal distribution will be remarked in Section 4 to be the well-known generating function for the Hermite orthogonal polynomials, with corresponding weight function provided by the central normal distribution. Determination of the entropic convex dual λ in terms of an expansion in a small number of low order Hermite polynomials will be carried through in Section 5. The translation factor for the noncentral ultraspherical F distribution will be shown in Section 6 to have an expansion in terms of Jacobi orthogonal polynomials, with corresponding weight function provided by the central ultraspherical F distribution. Determination of the entropic convex dual λ in terms of an expansion in a small number of low order Jacobi polynomials will be carried through in Section 7. Finally, the translation factor for the noncentral χ^2 distribution will be shown in Section 8 to have an expansion in terms of Laguerre orthogonal polynomials, with corresponding weight function provided by the central χ^2 distribution. Determination of the entropic convex dual λ in terms of an expansion in a small number of low order Laguerre polynomials will be carried through in Section 9.

Some Appendices are provided for completeness. Appendix A is devoted to the Mises-Fisher-Langevin ultraspherical distribution function together with its normalization factor, the normalized modified Bessel function of the first kind. The latter is provided with a novel expression for its Maclaurin expansion, revealing it as a generalization of the Maclaurin expansion for the hyperbolic cosine, and providing an exponential-like multiplicative factor in the expression of the translation factor $T^{\chi^2}_{\nu_1}$ for the classical noncentral χ^2 distribution. It is generalized in Appendix B to all four univariate noncentral distributions as classically derived, ultimately providing all of them with similar modular expressions of the form $T(r|r_o) = E(r|r_o)\ e^{-r_o^2/2}$ for their translation multiplicative factors, with $E(r|r_o)$ a generalized exponential, a factorization not suitable for expansion in terms of orthogonal polynomial families. Finally, the Kullback-Leibler divergence between the ultraspherical and the classical noncentral t and F distributions together with their respective conditions of applicability are discussed in Appendix C.

2. Noncentral ultraspherical t−distribution

The noncentral ultraspherical t-distribution for the t-statistic

$$t = \sqrt{\nu_2}\,\frac{\cos\theta}{\sin\theta}, \qquad 0 < \theta < \pi, \tag{5}$$

on the hypersphere $\mathcal{S}^{\nu_2}$ was shown by Le Blanc (2019) to be given by

$$\upsilon^t_{\nu_2}(\theta|\delta) = T^t_{\nu_2}(\theta|\delta)\ \upsilon^t_{\nu_2}(\theta|0), \tag{6}$$

where $\upsilon^t_{\nu_2}(\theta|0)$ stands for ultraspherical central t-distribution

$$\upsilon^t_{\nu_2}(\theta|0) = \frac{\Gamma(\frac{\nu_2+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu_2}{2})}\sin^{\nu_2-1}\theta, \tag{7}$$

where the multiplicative distribution translation factor $T^t_{\nu_2}(\theta|\delta)$ is given by

$$T^t_{\nu_2}(\theta|\delta) = \frac{1-\cos\theta\cos\theta_\delta}{[1-2\cos\theta\cos\theta_\delta+\cos^2\theta_\delta]^{(\nu_2+1)/2}}, \tag{8}$$

and where

$$\cos\theta_\delta = \frac{\delta/\sqrt{\nu_2}}{\sqrt{(\delta/\sqrt{\nu_2})^2+1}} = \frac{\delta}{\sqrt{\delta^2+\nu_2}}, \quad 0 < \theta_\delta < \pi, \tag{9}$$

in terms of the usual noncentrality parameter δ, $-\infty < \delta < \infty$. This compact expression for the noncentral t-distribution was obtained through a simple δ-specified translation of the hypersphere along its polar axis, with the translated distribution thereafter recomputed at the origin through a simple change of coordinates. The central distribution translation property of the multiplicative factor $T^t_{\nu_2}(\theta|\delta)$ will become obvious upon taking the limit $\nu_2 \to \infty$ below. Note that $T^t_{\nu_2}(\theta|0) = 1$, which ascertains that the noncentral ultraspherical t-distribution simplifies to the central ultraspherical t-distribution when the noncentrality parameter δ vanishes. As it turns out, the translation term $T^t_{\nu_2}(\theta|\delta)$ is a generating function for the classical ultraspherical or, equivalently, Gegenbauer polynomials. Indeed, redefining — for the sake of alignment on notational conventions of the extended literature on orthogonal polynomials — the noncentral ultraspherical t-distribution (6) variables so that $x = \cos\theta$, $z = \cos\theta_\delta$, and $b = (\nu_2-1)/2$, we have that

$$T^t_b(x|z) = \frac{1-xz}{[1-2xz+z^2]^{b+1}} = \sum_{n=0}^{\infty}\frac{2b+n}{2b}\,C^{(b)}_n(x)\,z^n,\ -1<x<1,\ -1<z<1, \tag{10}$$

where $C^{(b)}_n(x)$ are the Gegenbauer polynomials which can be provided with the explicit representation (Olver *et al.*, 2021)

$$C^{(b)}_n(x) = \sum_{k=0}^{\lfloor n/2\rfloor}\frac{(-1)^k(b)_{n-k}}{k!\,(n-2k)!}(2x)^{n-2k}, \tag{11}$$

with $\lfloor x \rfloor$ — the floor of x — given by the lowest integer such that $x-1 < \lfloor x \rfloor \leqq x$. The Gegenbauer polynomials are orthogonal with respect to the weight function

$$w_C^{(b)}(x) = (1-x^2)^{b-1/2}. \tag{12}$$

More precisely, we have that

$$\int_{z=-1}^{1} C_m^{(b)}(x)\; C_n^{(b)}(x)\; w_C^{(b)}(x)\; \mathrm{d}x = \delta_{m,n} \|C_n^{(b)}\|^2, \tag{13}$$

where

$$\|C_n^{(b)}\|^2 = \frac{\Gamma(\frac{1}{2})\Gamma(b+\frac{1}{2})}{\Gamma(b+1)}\; \frac{b}{b+n}\; \frac{(2b)_n}{n!}, \tag{14}$$

with $(x)_n$ the Pochhammer symbol defined by the equality

$$(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)} = x(x+1)(x+2)\cdots(x+n-1). \tag{15}$$

Equation (10) could be used to define a generalization $T_n^{(b)}(x)$ for the Chebyshev polynomials of the first kind, that is,

$$\frac{1-xz}{[1-2xz+z^2]^{b+1}} = \sum_{n=0}^{\infty} \frac{2b+n}{2b}\; C_n^{(b)}(x)\; z^n \equiv \sum_{n=0}^{\infty} T_n^{(b)}(z)\; z^n \tag{16}$$

which would encompass the generating function for the classical Chebyshev polynomials of the first kind with $b=0$, that is,

$$\frac{1-xz}{[1-2xz+z^2]} = \sum_{n=0}^{\infty} T_n^{(0)}(x)\; z^n = \lim_{b\to 0} \sum_{n=0}^{\infty} \frac{n}{2b}\; C_n^{(b)}(x)\; z^n. \tag{17}$$

Note the important fact that expansion (10) regroups all terms of same order in the noncentrality parameter z, a fact that we shall now exploit.

3. Gegenbauer polynomial expansion for the noncentral t distribution Gibbs prior

Consider dense datasets in high dimensional spaces. As extensively reviewed by Le Blanc (2022), Bayes-Jaynes-Gibbs data-constrained maximal entropy priors — simply designated Gibbs priors in the following — can be objectively computed for such dense datasets. The Gibbs priors can be expressed as

$$\pi(r_o) = \frac{1}{Z(\lambda)} \exp\left(\int_{\mathcal{R}} \lambda(r)\rho(r|r_o)\; \mathrm{d}r \right), \tag{18}$$

where the partition function $Z(\lambda)$ is given by

$$Z(\lambda) = \int_{\mathcal{R}_o} \exp\left(\int_{\mathcal{R}} \lambda(r)\rho(r|r_o)\; \mathrm{d}r \right) \mathrm{d}r_o, \tag{19}$$

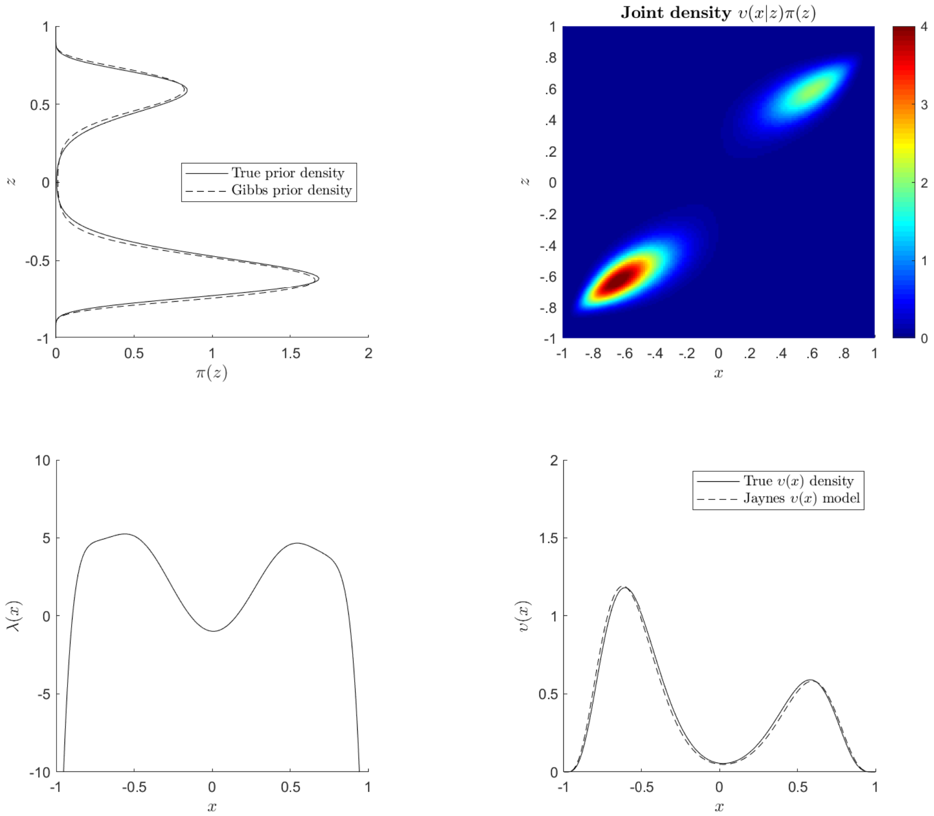

Figure 3.1: **Gibbs prior** for a mixture of noncentral ultrapherical t-distributions, with a target bimodal prior composed of the normalized sum of two such distributions. Left upper panel: Gibbs prior compared to the target prior, marginal density of the joint density to its right. Right upper panel: joint density $\upsilon(x|z)\pi(z)$. Left lower panel: Gibbs-Jaynes model entropic convex dual $\lambda(x)$. Right lower panel: Gibbs-Jaynes model density compared to the target density, marginal density of the joint density above it. The target density $\upsilon(x)$ has been provided here on its domain with a finite support of 200 points, but the optimization problem has been solved using a small 10-term Gegenbauer polynomial expansion for the entropic convex dual. See text for details.

and where $\lambda(r)$ is the entropic convex dual of the empirical density $\rho(r)$. It is obtained through the unconstrained minimization of the Gibbs potential

$$\inf_{\lambda} G_{\rho}(\lambda) = \inf_{\lambda} \left[\log Z(\lambda) - \int_{\mathcal{R}} \lambda(r)\rho(r) \, \mathrm{d}r \right], \tag{20}$$

which, by convex duality, corresponds to Jaynes data-constrained maximal entropy. The corresponding Gibbs-Jaynes model is thereafter given by

$$\rho(r) = \int_{\mathcal{R}_o} \rho(r|r_o) \; \pi(r_o) \; \mathrm{d}r_o. \tag{21}$$

As posed, solving for the entropic convex dual function $\lambda(r)$ in (20) requires one to compute its value across the entire support domain of the empirical distribution function $\rho(r)$, a task which can be expensive in computing terms. A much more economical computation of the Gibbs prior is possible for the four noncentral ultraspherical t, normal, ultraspherical F, and χ^2 distributions if one exploits the fact that their central distribution $\rho(r|0)$ are — to within a constant — the family-defining weight functions of the Gegenbauer, Hermite, Jacobi, and Laguerre orthogonal polynomial families, respectively, and that their translation factors $T(r|r_o)$ in the modular expression $\rho(r|r_o) = T(r|r_o)\rho(r|0)$ can be provided with an orthogonal polynomial expansion in terms of members of the respective families.

Since the noncentral ultraspherical t-distribution (6) is expressed as the product of a generating function (10) for the Gegenbauer polynomials times the corresponding weight function as provided by the central t-distribution (7), one can rewrite the unconstrained minimization problem (20) in simpler terms. With, as before, $x = \cos\theta$, $z = \cos\theta_\delta$, and $b = (\nu_2 - 1)/2$, the exponentiated term in the partition function (19) can be rewritten as

$$\begin{aligned} \int_{x=-1}^{1} \lambda(x) \; \upsilon(x|z) \; \mathrm{d}x &= \int_{x=-1}^{1} \lambda(x) \; \frac{1 - xz}{[1 - 2xz + z^2]^{b+1}} \; \frac{\Gamma(b+1)}{\Gamma(\frac{1}{2}) \; \Gamma(b+\frac{1}{2})} \; w_C^{(b)}(x) \; \mathrm{d}x \\ &= \sum_{n=0}^{\infty} \frac{\Gamma(b+1)}{\Gamma(\frac{1}{2}) \; \Gamma(b+\frac{1}{2})} \frac{2b+n}{2b} \left[\int_{x=-1}^{1} \lambda(x) \; C_n^{(b)}(x) \; w_C^{(b)}(x) \; \mathrm{d}x \right] z^n \\ &= \sum_{n=0}^{\infty} \frac{\Gamma(b+1)}{\Gamma(\frac{1}{2}) \; \Gamma(b+\frac{1}{2})} \; \frac{2b+n}{2b} \; \|C_n^{(b)}\|^2 \; \lambda_n^{(b)} \; z^n \\ &= \sum_{n=0}^{\infty} \frac{2b+n}{2(b+n)} \; \frac{(2b)_n}{n!} \; \lambda_n^{(b)} \; z^n \equiv \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(b)} \; z^n. \end{aligned} \tag{22}$$

Similarly, the additive term in (20) can be rewritten

$$\int_{x=-1}^{1} \lambda(x) \; \upsilon(x) \; \mathrm{d}x = \sum_{n=0}^{\infty} \; \lambda_n^{(b)} \int_{x=-1}^{1} C_n^{(b)}(x) \; \upsilon(x) \; \mathrm{d}x = \sum_{n=0}^{\infty} \; \tilde{\lambda}_n^{(b)} \tilde{\upsilon}_n^{(b)}, \tag{23}$$

where $\upsilon(x)$ is the empirical density to be modeled, and where

$$\tilde{\upsilon}_n^{(b)} = \frac{2(b+n)}{2b+n} \; \frac{n!}{(2b)_n} \; \int_{x=-1}^{1} C_n^{(b)}(x) \; \upsilon(x) \; \mathrm{d}x. \tag{24}$$

In particular, $\tilde{\upsilon}_o^{(b)} = 1$. Note that the latter integral does not involve the polynomial weight $w_C^{(b)}(x)$, so that one cannot ascribe the designation of orthogonal polynomial expansion term to the quantity $\tilde{\upsilon}_n^{(b)}$. We pause to remark on the fact that notational conventions regarding the Gegenbauer polynomials provide the 0^{th} order polynomial $C_o^{(b)}(x)$ with the non-unit norm

$$\|C_o^{(b)}\|^2 = \frac{\Gamma(\frac{1}{2})\Gamma(b+\frac{1}{2})}{\Gamma(b+1)}. \tag{25}$$

This stems from the historical fact that the weight function $w_C^{(b)}(x)$ for $b \geq 0$ is an unormalized ultraspherical central distribution. If the factor $1/\|C_o^{(b)}\|^2$ would have multiplied the weight function $w_C^{(b)}(x)$ therefore normalizing it, the norm of the constant polynomial $C_o^{(b)}$ would have been unity, and the above algebraic derivations would have been simplified. Conventions regarding normalization of the Hermite polynomials discussed in Section 4 involves a normalized weight function provided by the normal distribution $N(0,1)$ with, as a consequence, a simplified algebra. Conventions regarding the Jacobi polynomials in Section 6 and the Laguerre polynomials in Section 8 similarly involve non-normalized weight functions, which once more will encumber similar derivations. We shall depart from historical notational conventions in Section 8 concerning the Laguerre polynomials. The unconstrained optimization problem can thus be reformulated as

$$\inf_{\{\tilde{\lambda}_n^{(b)}\}} \left[\log\Big(\int_{z=-1}^{1} \exp(\sum_{n=0}^{\infty} \tilde{\lambda}_n^{(b)} z^n)\ \mathrm{d}z \Big) - \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(b)} \tilde{\upsilon}_n^{(b)} \right] \tag{26}$$

in terms of a Gegenbauer expansion $\{\lambda_n^{(b)}\}_{n=0}^{\infty}$ for the continuous entropic convex dual function $\lambda(x)$, expansion which can be restricted to a small number of coefficients. At the minimum of the Gibbs potential (26), one has the simple condition

$$\frac{\int_{z=-1}^{1} z^n \exp(\sum_n \tilde{\lambda}_n^{(b)} z^n)\ \mathrm{d}z}{\int_{z=-1}^{1} \exp(\sum_n \tilde{\lambda}_n^{(b)} z^n)\ \mathrm{d}z} = \tilde{\upsilon}_n^{(b)}, \tag{27}$$

which states that the n^{th} moment of the noncentrality parameter z is equal to the quantity provided by equality (24) when weighted by the Gibbs prior

$$\pi_C^{(b)}(z) = \frac{\exp(\sum_n \tilde{\lambda}_n^{(b)} z^n)}{\int_{z=-1}^{1} \exp(\sum_n \tilde{\lambda}_n^{(b)} z^n)\ \mathrm{d}z}. \tag{28}$$

Note that the constant (0^{th} order) expansion term coefficient $\tilde{\lambda}_o^{(b)}$ is trivially solved for, and need not be computed as it cancels itself out in the above ratio. In practical terms, determination of the Gibbs prior in terms of a small number of polynomial expansion coefficients $\tilde{\lambda}_n^{(b)}$ for the entropic dual convex results in a substantial reduction in computing time needed to find the Gibbs potential minimum. See Figure 3.1.

4. Noncentral normal distribution

It is known that if t distributes according to a noncentral t-distribution with degree of freedom ν_2 and noncentrality parameter δ, then $\lim_{\nu_2\to\infty} t$ should distribute according to a normal distribution with mean δ and unit variance. We shall verify this limit for the noncentral ultraspherical t-distribution (6). We begin by showing that the central ultraspherical t-distribution (7) tends to a normal distribution with null mean and unit variance. First, set $\theta = \frac{\pi}{2} - \phi$ in (7) to obtain

$$\upsilon^t_{\nu_2}(\phi|0) = \frac{\Gamma(\frac{\nu_2+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu_2}{2})} \cos^{\nu_2-1}\phi, \qquad -\frac{\pi}{2} < \phi < \frac{\pi}{2}, \tag{29}$$

transformation which will center the distribution on $\phi = 0$ rather than on $\theta = \pi/2$. Using the limit definition of the exponential function

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n, \tag{30}$$

and Stirling's approximation for the gamma function

$$\Gamma(z) \sim \sqrt{2\pi} e^{-z} z^{z-\frac{1}{2}}, \tag{31}$$

one obtains after a second change of variables $r^2 = \nu_2\phi^2$ that

$$\lim_{\nu_2\to\infty} \upsilon^t_{\nu_2}(r|0) = \rho^N(r|0) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}r^2}, \tag{32}$$

that is, the central ultraspherical t-distribution (7) tends to a normal distribution with null mean and unit variance, as expected. Using the same changes of variables and limiting procedures, the multiplicative factor (8) is shown to converge to the simple expression

$$\lim_{\nu_2\to\infty} T^t_{\nu_2}(r|\delta) = T^N(r|\delta) = e^{\delta r - \delta^2/2} = E^N(r|\delta)\, e^{-\delta^2/2} \tag{33}$$

where we introduce the notation

$$E^N(r|\delta) = e^{\delta r} = \sum_{j=0}^{\infty} \frac{(\delta r)^j}{j!} = {}_0F_0(;;\delta r) \tag{34}$$

for the exponential function e since generalizations of the exponential function will appear in a similar manner in the modular expressions of the other noncentral distributions. These generalizations can also be expressed in terms of the generalized hypergeometric function

$${}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_q;x) = \sum_{n=0}^{\infty} \frac{(a_1)_n\ldots(a_p)_n}{(b_1)_n\ldots(b_q)_n}\,\frac{x^n}{n!}. \tag{35}$$

The multiplicative factor $T^N(r|\delta)$ obviously imparts a non-vanishing mean δ to the above zero mean normal distribution since

$$\begin{aligned} T^N(r|\delta)\times\rho^N(r|0) &= e^{\delta r-\delta^2/2}\times\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}r^2} \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(r-\delta)^2} = \rho^N(r|\delta). \end{aligned} \tag{36}$$

The orthogonal polynomial literature allows us to retrieve the same limit results. Indeed, the multiplicative term $T^N(r|\delta) = e^{\delta r - \delta^2/2}$ can be recognized as the generating function for the Hermite polynomials, that is,

$$T^N(r|\delta) = e^{r\delta - \delta^2/2} = \sum_{n=0}^{\infty} He_n(r)\ \frac{\delta^n}{n!}, \tag{37}$$

with

$$He_n(r) = n! \sum_{\ell=0}^{\lfloor n/2 \rfloor} \frac{(-1)^\ell\ r^{n-2\ell}}{2^\ell \ell!\ (n-2\ell)!} \tag{38}$$

an explicit representation for the Hermite polynomials (Olver *et al.*, 2021). The Hermite polynomials are orthogonal with respect to their defining weight function $w_H(r) = \rho^N(r|0)$, that is, the central normal $N(0,1)$ distribution. We have that

$$\int_{r=-\infty}^{\infty} He_m(r)\ He_n(r)\ \rho^N(r|0)\ \mathrm{d}r = \delta_{m,n}\ n!\ . \tag{39}$$

Applying the limit $\nu_2 \to \infty$ to the Gegenbauer polynomial expansion (10) for the translation factor $T_b^t(x|z)$, we find, with $b = (\nu_2 - 1)/2$, $x = r/(2b)^{1/2}$ and $z = \delta/(2b)^{1/2}$, that

$$\begin{aligned}
\lim_{b\to\infty} T_b^t(x|z) &= \lim_{b\to\infty} \frac{1 - xz}{[1 - 2xz + z^2]^{b+1}} \\
&= \lim_{b\to\infty} \sum_{n=0}^{\infty} \frac{2b+n}{2b}\ C_n^{(b)}\Big(\frac{r}{(2b)^{1/2}}\Big)\ \Big(\frac{\delta}{(2b)^{1/2}}\Big)^n \\
&= \sum_{n=0}^{\infty} He_n(r) \frac{\delta^n}{n!} = e^{r\delta - \delta^2/2} = T^N(r|\delta),
\end{aligned} \tag{40}$$

where we have used the limit result (Olver *et al.*, 2021)

$$\lim_{b\to\infty} \frac{1}{(2b)^{n/2}}\ C_n^{(b)}\Big(\frac{r}{(2b)^{1/2}}\Big) = \frac{He_n(r)}{n!}. \tag{41}$$

Note that expansion (37) regroups all terms of same order in the noncentrality parameter δ, a fact that we shall now exploit.

5. Hermite polynomial expansion for the noncentral normal distribution Gibbs prior

The simplified approach to the determination of the Gibbs prior for the noncentral ultraspherical t distribution in terms of a Gegenbauer polynomial expansion as outlined in Section 3 can be carried through step by step when applied to the computation of the Gibbs prior for the noncentral normal distribution in terms of a Hermite polynomial expansion. Recall that the noncentral distribution translation factor $T^N(r|\delta)$ provides the generating function (37) for the Hermite polynomials.

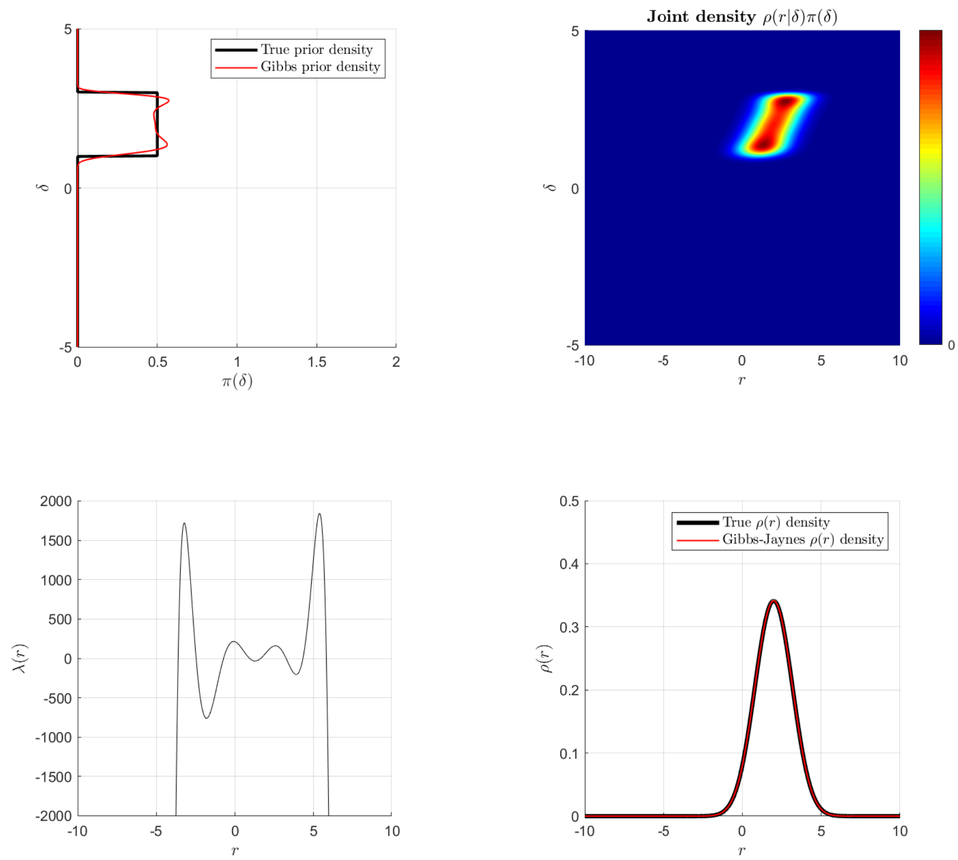

Figure 5.1: **Gibbs prior approximation** for a target Bayesian prior defined by a uniform density with sharp square boundaries. Using a finite expansion on the first eight Hermite polynomials for the entropic dual convex function $\lambda(r)$, the Gibbs prior reproduces the target square prior fairly accurately even though its input, the target empirical density $\rho(r)$, does not betray the square shape of its underlying prior. The expected oscillatory Gibbs phenomenon is observed at the boundaries of the Gibbs prior.

Consequently, the exponentiated term in the partition function (19) for the Gibbs potential can be rewritten in terms of the Hermite polynomials as

$$\begin{aligned}\int_{r=-\infty}^{\infty} \lambda(r)\ \rho(r|\delta)\ \mathrm{d}r &= \int_{r=-\infty}^{\infty} \lambda(r)\ e^{r\delta-\delta^2/2}\rho(r|0)\ \mathrm{d}r \\ &= \sum_{n=0}^{\infty}\left[\int_{r=-\infty}^{\infty} \lambda(r)\ He_n(r)\ \rho(r|0)\ \mathrm{d}r\right]\ \frac{\delta^n}{n!} \\ &= \sum_{n=0}^{\infty} \lambda_n\ \delta^n, \qquad (42)\end{aligned}$$

where the Hermite expansion coefficient λ_n for the entropic convex dual $\lambda(r)$ is given by

$$\lambda_n = \frac{1}{n!}\int_{r=-\infty}^{\infty} \lambda(r)\ He_n(r)\ \rho(r|0)\ \mathrm{d}r. \qquad (43)$$

Similarly, the additive term in (20) can be rewritten

$$\int_{r=-\infty}^{\infty} \lambda(r)\ \rho(r)\ \mathrm{d}r = \sum_{n=0}^{\infty}\ \lambda_n\ \rho_n \qquad (44)$$

where $\rho(r)$ is the empirical density to be modeled, and where

$$\rho_n = \int_{r=-\infty}^{\infty} He_n(r)\ \rho(r)\ \mathrm{d}r. \qquad (45)$$

In particular, $\rho_o = 1$. The unconstrained optimization problem can be reformulated as

$$\inf_{\{\lambda_n\}}\left[\log\Big(\int_{\delta=-\infty}^{\infty} \exp(\sum_{n=0}^{\infty}\lambda_n\ \delta^n)\ \mathrm{d}\delta\Big) - \sum_{n=0}^{\infty}\lambda_n\delta^n\right] \qquad (46)$$

in terms of a Hermite expansion $\{\lambda_n\}$ for the continuous entropic convex dual function $\lambda(r)$. At the minimum of the Gibbs potential (26), one has the simple condition

$$\frac{\int_{\delta=-\infty}^{\infty} \delta^n\ \exp(\sum_{n=0}^{\infty}\lambda_n\ \delta^n)\ \mathrm{d}\delta}{\int_{\delta=-\infty}^{\infty} \exp(\sum_{n=0}^{\infty}\lambda_n\ \delta^n)\ \mathrm{d}\delta} = \rho_n \qquad (47)$$

which states that the n^{th} moment of the noncentrality parameter δ is equal to the quantity ρ_n provided by the equality (45) when weighted by the Gibbs prior

$$\pi_H(\delta) = \frac{\exp(\sum_{n=0}^{\infty}\lambda_n\ \delta^n)}{\int_{\delta=-\infty}^{\infty} \exp(\sum_{n=0}^{\infty}\lambda_n\ \delta^n)\ \mathrm{d}\delta}. \qquad (48)$$

Note that the constant (0^{th} order) expansion term λ_o is trivially solved for, and need not be computed as it cancels itself out in the above ratio. The Hermite expansion for the dual conjugate $\lambda(r)$ can be restricted to a small number of low order Hermite polynomials. See Figure 5.1 for a challenging case involving a target square Bayes prior, with its Gibbs prior being affected by an expected oscillatory Gibbs phenomenon (Shizgal & Jung, 2003).

6. Noncentral ultraspherical $F-$distribution

The noncentral ultraspherical F-distribution for the F-statistic

$$F = \frac{\nu_2}{\nu_1}\frac{\cos^2\theta}{\sin^2\theta}, \qquad 0 \leq \theta \leq \frac{\pi}{2}, \tag{49}$$

on the hypersphere was shown by Le Blanc (2019) to be given by the integral representation

$$\upsilon^F_{(\nu_1,\nu_2)}(\theta|\Lambda) = T^F_{(\nu_1,\nu_2)}(\theta|\Lambda)\ \upsilon^F_{(\nu_1,\nu_2)}(\theta|0), \tag{50}$$

where $\upsilon^F_{(\nu_1,\nu_2)}(\theta|0)$ stands for central hypersphere F-distribution

$$\upsilon^F_{(\nu_1,\nu_2)}(\theta|0) = 2\frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\ \cos^{\nu_1-1}\theta\ \sin^{\nu_2-1}\theta, \tag{51}$$

where the multiplicative distribution translation factor $T^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ is given by the integral

$$T^F_{(\nu_1,\nu_2)}(\theta|\Lambda) = \int_{\psi=0}^{\pi} T^F_{(\nu_1,\nu_2)}(\theta,\psi|\Lambda)\ \mathrm{d}\psi \tag{52}$$

with integrand

$$T^F_{(\nu_1,\nu_2)}(\theta,\psi|\Lambda) = \frac{(1-\cos\theta\cos\psi\cos\theta_\Lambda)}{[1-2\cos\theta\cos\psi\cos\theta_\Lambda+\cos^2\theta_\Lambda]^{(\nu_1+\nu_2)/2}}\ \upsilon^t_{\nu_1-1}(\psi|0), \tag{53}$$

and where

$$\cos\theta_\Lambda = \sqrt{\Lambda/(\Lambda+\nu_2)}, \qquad 0 \leq \Lambda < \infty, \quad 0 \leq \cos\theta_\Lambda < 1, \tag{54}$$

in terms of the noncentral parameter Λ. The appearance of the ψ coordinate in the integrand $T^F_{(\nu_1,\nu_2)}(\theta,\psi|\Lambda)$ stems from the fact that the noncentral ultraspherical F-distribution is obtained by translating the unit radius central hypersphere along a Λ-specified vector lying on a secant ν_1-dimensional hyperplane. Any normalized vector stemming from this translated hypersphere can be described in the original coordinate system by a normalized three-vector $(\cos\theta\cos\psi, \cos\theta\sin\psi, \sin\theta)$, with entries referring to the vector's projections along the Λ-specified translation axis, its orthogonal complement in $\mathcal{S}^{\nu_1}$, and its orthogonal complement in $\mathcal{S}^{\nu_2}$, respectively. This three-vector distributes according to the probability distribution

$$\upsilon^F_{(\nu_1,\nu_2)}(\theta,\psi|\Lambda) = T^F_{(\nu_1,\nu_2)}(\theta,\psi|\Lambda)\ \upsilon^F_{(\nu_1,\nu_2)}(\theta|0), \quad 0 \leq \theta \leq \pi/2, \quad 0 \leq \psi \leq \pi, \tag{55}$$

which provides us with the integrand for the integral representation of the noncentral F-distribution above. When $\Lambda = 0$, we have that $\cos\theta_\Lambda = 0$ and that

$$T^F_{(\nu_1,\nu_2)}(\theta|\Lambda=0) = \int_{\psi=0}^{\pi} T^F_{(\nu_1,\nu_2)}(\theta,\psi|\Lambda=0)\ \mathrm{d}\psi = \int_{\psi=0}^{\pi} \upsilon^t_{\nu_1-1}(\psi|0)\ \mathrm{d}\psi = 1, \tag{56}$$

which ascertains that the noncentral ultraspherical F-distribution $\upsilon^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ simplifies to the central ultraspherical F-distribution $\rho^F_{(\nu_1,\nu_2)}(\theta|0)$ when the noncentral parameter Λ vanishes. The special case $\nu_1 = 1$ is given by

$$\upsilon^F_{(\nu_1=1,\nu_2)}(\theta|\Lambda) = \sum_{\cos\theta'=[\cos\theta,-\cos\theta]} \frac{1-\cos\theta'\cos\theta_\Lambda}{[1-2\cos\theta'\cos\theta_\Lambda+\cos^2\theta_\Lambda]^{(\nu_2+1)/2}}\ \upsilon^t_{\nu_2}(\theta|0). \tag{57}$$

One can explicitly compute an orthogonal polynomial expansion for the translating factor $T^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ by exploiting the Gegenbauer polynomial generating function (10). Setting

$$\begin{aligned} a &= (\nu_1-2)/2,\ b=\nu_2/2, \\ x &= \cos\theta,\ \xi=\cos\psi,\ z=\cos\theta_\Lambda \end{aligned} \tag{58}$$

the integration in equation (52) can be analytically performed to get after some algebra

$$\begin{aligned} T^F_{(a,b)}(x|z) &= \int_{\xi=-1}^{1} \frac{1-x\xi z}{[1-2x\xi z+z^2]^{a+b+1}} \frac{\Gamma(a+1)}{\Gamma(\frac{1}{2})\Gamma(\frac{a+1}{2})}\ w_C^{(a)}(\xi)\ \mathrm{d}\xi \\ &= \sum_{n=0}^{\infty} \frac{a+b+n}{a+b}\ F_n^{(a,b)}(x)\ z^{2n}, \quad 0<z<1, \end{aligned} \tag{59}$$

where the $F_n^{(a,b)}(x)$ polynomials are provided with the explicit representation

$$F_n^{(a,b)}(x) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \frac{(a+b)_{2n-k}}{(a+1)_{n-k}} \frac{(x^2)^{n-k}}{(n-k)!}, \quad 0<x<1. \tag{60}$$

That the $F_n^{(a,b)}(x)$ polynomials are orthogonal

$$\int_{x=0}^{1} F_m^{(a,b)}(x)\ F_n^{(a,b)}(x)\ w_F^{(a,b)}(x)\ \mathrm{d}x = \delta_{m,n}\|F_n^{(a,b)}\|^2 \tag{61}$$

with respect to the weight function

$$w_F^{(a,b)}(x) = (x^2)^{a+\frac{1}{2}}\ (1-x^2)^{b-1}, \quad 0<x<1, \tag{62}$$

stems from the fact that, under the change of variable $x^2=(1-y)/2$, the weight function $w_F^{(a,b)}(x)$ becomes proportional to the weight function

$$w_P^{(\alpha,\beta)} = (1-y)^{\alpha}\ (1+y)^{\beta}, \quad -1<y<1, \tag{63}$$

for the Jacobi polynomials $P_n^{(\alpha,\beta)}(y)$, $\alpha=a$, $\beta=b-1$, and from the fact that using the explicit representation

$$P_n^{(\alpha,\beta)}(y) = \sum_{\ell=0}^{\infty} \frac{(\alpha+\beta+1+n)_\ell(\alpha+\ell+1)_{n-\ell}}{\ell!\ (n-\ell)!} \left(\frac{y-1}{2}\right)^{\ell}, \tag{64}$$

for the Jacobi polynomials (Olver *et al.*, 2021), one readily verifies that

$$F_n^{(a,b)}(x) = (-1)^n \frac{(a+b)_n}{(a+1)_n}\ P_n^{(a,b-1)}(1-2x^2), \tag{65}$$

with norm

$$\|F_n^{(a,b)}\|^2 = \frac{a+b}{a+b+2n} \frac{(a+b)_n(b)_n}{n!(a+1)_n} \times \frac{\Gamma(a+1)\Gamma(b)}{2\Gamma(a+b+1)}, \tag{66}$$

where the last ratio is the inverse of the normalization factor for the central F distribution (51). If the weight $w_F^{(a,b)}(x)$ would have been multiplied by this normalization factor, the last ratio in the norm of F_n would have disappeared, and the norm of the 0^{th} order polynomial would have simplified to $\|F_o^{(a,b)}\|^2 = 1$. The polynomials $P_n^{(a,b-1)}(1-2x^2)$ will be argued in Section 8 to converge unto the Laguerre polynomials $L_n^{(a)}$ in the limit $b \to \infty$. The polynomial family $\{F_n^{(a,b)}(x)\}_{n=0}^{\infty}$ thus possesses the orthogonality and completeness properties of the Jacobi polynomials. Note that expansion (59) regroups all terms of same order in the noncentrality parameter z, a fact that we shall now exploit.

7. Jacobi polynomial expansion for the noncentral F distribution Gibbs prior

Since the noncentral F-distribution can be expressed as the product of the expansion (59) in terms of the orthogonal polynomials $F_n^{(a,b)}(x)$, times the corresponding weight function $w_F^{(a,b)}(x)$ as provided by equation (62), one can rewrite the unconstrained minimization problem (20) in simpler terms. The following development can be compared to that outlined in Section 3. The exponentiated term in the partition function (19) can be rewritten as

$$\begin{aligned}\int_{x=0}^{1} \lambda(x)\upsilon(x|z)\mathrm{d}x &= 2\,\frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} \sum_{n=0}^{\infty} \frac{a+b+n}{a+b} \left[\int_{x=0}^{1} \lambda(x)\ F_n^{(a,b)}(x)\ w_F^{(a,b)}(x)\ \mathrm{d}x\right] z^n \\ &= 2\,\frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} \sum_{n=0}^{\infty} \frac{a+b+n}{a+b}\ \|F_n^{(a,b)}\|^2\ \lambda_n^{(a,b)}\ z^n \\ &= \sum_{n=0}^{\infty} \frac{a+b+n}{a+b+2n}\ \frac{(a+b)_n(b)_n}{n!(a+1)_n}\ \lambda_n^{(a,b)}\ z^n \equiv \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a,b)}\ z^n. \end{aligned} \tag{67}$$

As in Section 3, we remark that the preceding expressions would have been simplified if the terms in front of the summation signs would have been used to normalize the weight function $w_F^{(a,b)}$ to the central ultraspherical F distribution (51). In a similar fashion, the additive term in (20) can be rewritten

$$\int_{x=0}^{1} \lambda(x)\ \upsilon(x)\ \mathrm{d}x = \sum_{n=0}^{\infty} \lambda_n^{(a,b)} \int_{x=0}^{1} F_n^{(a,b)}(x)\ \upsilon(x)\ \mathrm{d}x = \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a,b)}\tilde{\upsilon}_n^{(a,b)}, \tag{68}$$

where $\upsilon(x)$ is the empirical density to be modeled, and where

$$\tilde{\upsilon}_n^{(a,b)} = \frac{a+b+2n}{a+b+n}\ \frac{n!(a+1)_n}{(a+b)_n(b)_n} \int_{x=0}^{1} F_n^{(a,b)}(x)\ \upsilon(x)\ \mathrm{d}x. \tag{69}$$

In particular, $\tilde{\upsilon}_o^{(a,b)} = 1$. The unconstrained optimization problem can than be reformulated as

$$\inf_{\{\tilde{\lambda}_n^{(a,b)}\}} \left[\log\Big(\int_{z=0}^{1} \exp(\sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a,b)} z^n)\ \mathrm{d}z\Big) - \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a,b)}\tilde{\upsilon}_n^{(a,b)}\right] \tag{70}$$

in terms of a Jacobi expansion $\{\lambda_n^{(a,b)}\}_{n=0}^{\infty}$ for the continuous entropic convex dual function $\lambda(x)$, expansion which can be restricted to a small number of coefficients. At the minimum of the Gibbs potential (26), one has the simple condition

$$\frac{\int_{z=0}^{1} z^n \exp(\sum_n \tilde{\lambda}_n^{(a,b)} z^n) \, \mathrm{d}z}{\int_{z=0}^{1} \exp(\sum_n \tilde{\lambda}_n^{(a,b)} z^n) \, \mathrm{d}z} = \tilde{\upsilon}_n^{(a,b)}, \tag{71}$$

which states that the n^{th} moment of the noncentrality parameter z is equal to the quantity provided by equality (69) when weighted by the Gibbs prior

$$\pi_F^{(a,b)}(z) = \frac{\exp(\sum_n \tilde{\lambda}_n^{(a,b)} z^n)}{\int_{z=0}^{1} \exp(\sum_n \tilde{\lambda}_n^{(a,b)} z^n) \, \mathrm{d}z}. \tag{72}$$

Note that the constant (0^{th} order) expansion term $\tilde{\lambda}_o^{(a,b)}$ is trivially solved for, and need not be computed as it cancels itself out in the above ratio.

8. Noncentral χ^2 distribution

It is known that if F distributes according to a noncentral F-distribution with degrees of freedom (ν_1, ν_2) and noncentrality parameter Λ, then $\lim_{\nu_2 \to \infty} \nu_1 F$ should distribute according to a noncentral χ^2 distribution with noncentrality parameter Λ. We proceed with this limit for the integral representation of the noncentral ultraspherical F-distribution (50) in order to derive an integral representation for the noncentral χ^2 distribution. We begin by showing that the central ultraspherical F-distribution (51) tends to a central $\chi^2_{\nu_1}$ distribution. First, set $\theta = \frac{\pi}{2} - \phi$ in (51) to obtain

$$\upsilon^F_{(\nu_1,\nu_2)}(\phi|0) = 2\frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \sin^{\nu_1-1}\phi \cos^{\nu_2-1}\phi, \qquad 0 \le \phi \le \frac{\pi}{2}. \tag{73}$$

Using once again the limit definition of the exponential function (30) and Stirling's approximation for the gamma function (31), one obtains after a second change of variables $r = \nu_2 \phi^2$ that

$$\lim_{\nu_2 \to \infty} \upsilon^F_{(\nu_1,\nu_2)}(r|0) = \frac{1}{2^{\nu_1/2}\Gamma(\frac{\nu_1}{2})} r^{(\nu_1-2)/2} e^{-r/2} = \rho^{\chi^2}_{\nu_1}(r|0), \quad 0 \le r < \infty, \tag{74}$$

in terms of the central $\chi^2_{\nu_1}$ distribution $\rho^{\chi^2}_{\nu_1}(r|0)$. Using the same changes of variables and limiting procedures, the multiplicative distribution translation factor (53) is shown to converge to the simple expression

$$\begin{aligned} \lim_{\nu_2 \to \infty} T^F_{(\nu_1,\nu_2)}(r,\psi|\Lambda) &= T^{\chi^2}_{\nu_1}(r,\psi|\Lambda), \qquad 0 \le \psi \le \pi, \\ &= e^{\sqrt{\Lambda r}\cos\psi - \Lambda/2} \; \upsilon^t_{\nu_1-1}(\psi|0) \\ &= \mathcal{I}_{(\nu_1-2)/2}(\Lambda r) \; e^{-\Lambda/2} \; \upsilon^{\text{vMFL}}_{\nu_1-1}(\psi|\sqrt{\Lambda r}) \end{aligned} \tag{75}$$

where we have used definition (96) for the vMFL ultraspherical distribution with its normalization factor (98) as provided by the normalized modified Bessel function

$\mathcal{I}_{(\nu_1-2)/2}(\Lambda r)$. Comparing the above translation factor $T^{\chi^2}_{\nu_1}(r,\psi|\Lambda)$ to the translation factor $T^N(r|\delta)$ in (33), and observing that the parameters of a χ^2 distribution correspond to the square of the corresponding parameters of a normal distribution, it is seen that the term

$$E^{\chi^2}_{\nu_1}(r|\Lambda) = \mathcal{I}_{(\nu_1-2)/2}(\Lambda r) = {}_0F_1\left(;\frac{\nu_1}{2};\frac{\Lambda r}{4}\right) \tag{76}$$

corresponds to the term $e^{\delta r}$, and the term $e^{-\Lambda/2}$ to the term $e^{-\delta^2/2}$, respectively. The integrand for the integral representation of the noncentral $\chi^2_{\nu_1}$ distribution is thus given by

$$\rho^{\chi^2}_{\nu_1}(r,\psi|\Lambda) = E^{\chi^2}_{\nu_1}(r|\Lambda)\; e^{-\Lambda/2}\; \rho^{\chi^2}_{\nu_1}(r|0)\; \upsilon^{\mathrm{vMFL}}_{\nu_1-1}(\psi|\,\sqrt{\Lambda r}). \tag{77}$$

Performing the integral over the ψ variable, one obtains

$$\begin{aligned}\rho^{\chi^2}_{\nu_1}(r|\Lambda) &= \int_{\psi=0}^{\pi} \rho^{\chi^2}_{\nu_1}(r,\psi|\Lambda)\; \mathrm{d}\psi \\ &= E^{\chi^2}_{\nu_1}(r|\Lambda)\; e^{-\Lambda/2}\; \rho^{\chi^2}_{\nu_1}(r|0) \int_{\psi=0}^{\pi} \upsilon^{\mathrm{vMFL}}_{\nu_1-1}(\psi|\,\sqrt{\Lambda r})\; \mathrm{d}\psi \\ &= E^{\chi^2}_{\nu_1}(r|\Lambda)\; e^{-\Lambda/2}\; \rho^{\chi^2}_{\nu_1}(r|0). \end{aligned} \tag{78}$$

The noncentral $\chi^2_{\nu_1}$ distribution is thus given as the product of the central $\chi^2_{\nu_1}$ distribution times the translation factor

$$T^{\chi^2}_{\nu_1}(r|\Lambda) = E^{\chi^2}_{\nu_1}(r|\Lambda)\; e^{-\Lambda/2} = \mathcal{I}_{(\nu_1-2)/2}(\Lambda r)\; e^{-\Lambda/2} = {}_0F_1\left(;\frac{\nu_1}{2};\frac{\Lambda r}{4}\right)\; e^{-\Lambda/2}. \tag{79}$$

Using the limiting results for the noncentral t-distribution in Section 4, the limiting procedure for the special case $\upsilon^F_{(\nu_1=1,\nu_2)}(\theta|\Lambda)$ as provided by equation (57) gives

$$T^{\chi^2}_{\nu_1=1}(r|\Lambda) = \cosh\sqrt{\Lambda r}\; e^{-\Lambda/2} = \mathcal{I}_{-1/2}(\Lambda r)\; e^{-\Lambda/2} = {}_0F_1\left(;\frac{1}{2};\frac{\Lambda r}{4}\right)\; e^{-\Lambda/2}, \tag{80}$$

that is, equation (79) is shown to be valid for all cases $\nu_1 \geq 1$. Since $E^{\chi^2}_{\nu_1}(r|0) = 1$ for all values of ν_1, the noncentral $\chi^2_{\nu_1}(r|\Lambda)$ distribution is ascertained to simplify to the central $\chi^2_{\nu_1}(r|0)$ distribution when the noncentral parameter Λ vanishes.

The classical expression (78) for the noncentral χ^2 distribution unfortunately does not regroup term of similar orders in the noncentral parameter Λ. As a consequence, an expansion in terms of orthogonal polynomials is not possible with the latter. Now, we have seen in Section (4) that the noncentral t distribution (6) converges unto that of a noncentral normal distribution (36) in the limit of large samples, and that the generating function for the Gegenbauer polynomials (10) converges accordingly unto the generating function for the Hermite polynomials (37). A similar correspondence holds for the noncentral F and χ^2 distributions. First, note that the central χ^2 distribution $\rho^{\chi^2}_{\nu_1}(r|0)$ as given by equation (74) is proportional to the weight function $x^\alpha e^{-x}$ of the orthogonal Laguerre polynomial family, with here

$x = r/2$ and $\alpha = a = (\nu_1 - 2)/2$. We choose to depart in this last example with the usual conventions of the orthogonal polynomial literature by using the normalized central χ^2 distribution

$$w_L^{(a)}(r) = \frac{1}{2^{a+1}} \frac{1}{\Gamma(a+1)} r^a e^{-r/2}, \quad a > -1, \tag{81}$$

as weight function for the Laguerre polynomials $L_n^{(a)}(y)$ which can be given the explicit representation (Olver *et al.*, 2021)

$$L_n^{(a)}(y) = \sum_{\ell=0}^{n} (-1)^\ell \, \frac{(\alpha + \ell + 1)_{n-\ell}}{(n-\ell)!} \, \frac{y^\ell}{\ell!}, \quad 0 \le y < \infty. \tag{82}$$

Note then that equations (59) and (65) provide the noncentral F distribution translation factor $T^F_{(a,b)}(x|z)$ with an expansion in terms of the Jacobi polynomials $P^{(a,b-1)}(1-2x^2)$. Since the limit result

$$\lim_{\beta\to\infty} P_n^{(\alpha,\beta)}\left(1 - \left(\frac{2y}{\beta}\right)\right) = L_n^{(\alpha)}(y) \tag{83}$$

holds (Olver *et al.*, 2021), one readily verifies that with $x^2 = r/2b$ and $z^2 = \Lambda/2b$

$$\lim_{b\to\infty} T^F_{(a,b)}(x|z) = T_a^{\chi^2}(r|\Lambda) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n (a+1)_n} \, L_n^{(a)}(r/2) \, \Lambda^n, \tag{84}$$

a lesser known expression for the noncentral χ^2 distribution first derived by Tiku (1965). With our convention regarding the weight function (89), the norm of the Laguerre polynomials $L_n^{(a)}(r/2)$ is given by

$$\int_{r=0}^{\infty} L_m^{(a)}(r/2) \, L_n^{(a)}(r/2) \, w_L^{(a)}(r) \, \mathrm{d}r = \|L_n^{(a)}\|^2 \, \delta_{m,n} = \frac{(a+1)_n}{n!} \, \delta_{m,n}. \tag{85}$$

Note that expansion (84) regroups all terms of same order in the noncentrality parameter Λ, a fact that we shall now exploit.

9. Laguerre polynomial expansion for the noncentral χ^2 distribution Gibbs prior

Since the noncentral χ^2 distribution can be expressed as the product of expansion (84) in terms of the orthogonal Laguerre polynomials $L_n^{(a)}(r/2)$, times the corresponding weight function $w_L^{(a)}(r/2)$ as provided by equation (81), one can rewrite the unconstrained minimization problem (20) in simpler terms. The exponentiated term in the partition function (19) can be rewritten as

$$\begin{aligned}
\int_{r=0}^{\infty} \lambda(r) \, T_{(a)}^{\chi^2}(r|\Lambda) \, w_L^{(a)}(r) \, \mathrm{d}r &= \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n (a+1)_n} \left[\int_{r=0}^{\infty} \lambda(r) \, L_n^{(\alpha)}(r/2) \, w_L^{(a)}(r) \, \mathrm{d}r\right] \Lambda^n \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n (a+1)_n} \lambda_n^{(a)} \|L_n^{(a)}\|^2 \, \Lambda^n \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} \, \lambda_n^{(a)} \Lambda^n \equiv \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \, \Lambda^n.
\end{aligned} \tag{86}$$

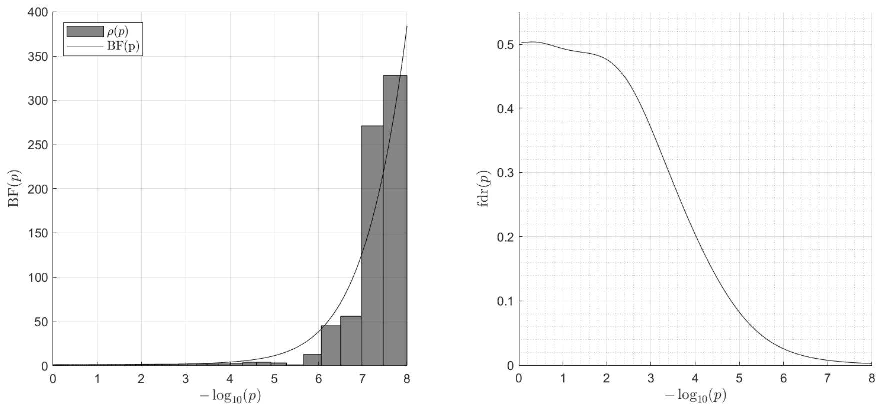

Figure 9.1: **Bayes Factor** $BF(p)$ modeling a NHST p-value distribution from a genome-wide association study (GWAS) dataset. The GWAS compared 2,244 critically ill patients with COVID-19 with three times as many ancestry-matched control individuals. The dataset comprises 4,380,209 $\chi^2_{\nu_1=1}$ r-statistics accounting for all the SNPs in the set which have been modelized by a logistic regression model and tested for statistical significance. Left panel: The NHST p-value empirical density $\rho(p(r))$ is seen to be well-approximated by the Bayes Factor $BF(p(r)) = \int_\Lambda T^{\chi^2}_{-\frac{1}{2}}(p(r)|\Lambda)\ \pi_L^{\left(-\frac{1}{2}\right)}(\Lambda)\ \mathrm{d}\Lambda$, where the Gibbs prior $\pi_L^{\left(-\frac{1}{2}\right)}(\Lambda)$ is provided in terms of an entropic dual convex function $\lambda(r)$ expanded on only six low order Laguerre polynomials $\{L_n^{\left(-\frac{1}{2}\right)}(r/2)\}_{n=1}^{6}$. See text for details. Right panel: Local false discovery rate $\mathrm{fdr}(p) = 1/(1+BF(p))$. The fdr crosses the 0.01 threshold (i.e. a fdr of 1%) when the NHST p-value reaches about 10^{-7}, that is $-\log_{10}(p) = 7$, in close concordance with the threshold of significance of 5×10^{-8}, that is $-\log_{10}(p) = 7.3$, chosen by the authors.

Similarly, the additive term in (20) can be rewritten

$$\int_{r=0}^{\infty} \lambda(r)\ \rho(r)\ \mathrm{d}r = \sum_{n=0}^{\infty} \lambda_n^{(a)} \int_{r=0}^{\infty} L_n^{(a)}(r/2)\ \rho(r)\ \mathrm{d}r = \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \tilde{\rho}_n^{(a)}, \tag{87}$$

where $\rho(r)$ is the empirical density to be modeled, and where

$$\tilde{\rho}_n^{(a)} = (-2)^n\ n! \int_{r=0}^{\infty} L_n^{(a)}(r/2)\ \rho(r)\ \mathrm{d}r. \tag{88}$$

In particular, $\tilde{\rho}_o^{(a)} = 1$. The unconstrained optimization problem can than be reformulated as

$$\inf_{\{\tilde{\lambda}_n^{(a)}\}} \left[\log\Big(\int_{\Lambda=0}^{\infty} \exp\left(\sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \Lambda^n \right)\ \mathrm{d}\Lambda \Big) - \sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \tilde{\rho}_n^{(a)} \right] \tag{89}$$

in terms of a Laguerre expansion $\{\lambda_n^{(a)}\}_{n=0}^{\infty}$ for the continuous entropic convex dual function $\lambda(r)$, expansion which can be restricted to a small number of coefficients. At the minimum of the Gibbs potential (26), one has the simple condition

$$\frac{\int_{\Lambda=0}^{\infty} \Lambda^n \exp\left(\sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \Lambda^n \right)\ \mathrm{d}\Lambda}{\int_{\Lambda=0}^{\infty} \exp\left(\sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \Lambda^n \right)\ \mathrm{d}\Lambda} = \tilde{\rho}_n^{(a)}, \tag{90}$$

which states that the n^{th} moment of the noncentrality parameter Λ is equal to the quantity provided by equality (88) when weighted by the Gibbs prior

$$\pi_L^{(a)}(\Lambda) = \frac{\exp\left(\sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \Lambda^n \right)}{\int_{\Lambda=0}^{\infty} \exp\left(\sum_{n=0}^{\infty} \tilde{\lambda}_n^{(a)} \Lambda^n \right)\ \mathrm{d}\Lambda}. \tag{91}$$

Note that the constant (0^{th} order) expansion term $\tilde{\lambda}_o^{(a)}$ is trivially solved for, and need not be computed as it cancels itself out in the above ratio.

We close this section by illustrating the strengths of the present approach in analyzing a contemporary genomic-scale dataset, with the strong caveat that this exercise should be considered only as a proof of principle rather than as a definitive approach to genome-wide association study (GWAS) dataset analysis. A GWAS is an observational study assessing a genome-wide set of genetic variants in different individuals, and seeking to identify statistically significant variant-trait associations. GWAS commonly focus on associations between single-nucleotide polymorphisms (SNPs) and traits. We retrieved the GenoMICC EUR vs UK biobank controls dataset from the GenOMICC (Genetics Of Mortality In Critical Care) GWAS comparing 2,244 critically ill patients with COVID-19 from UK intensive care units with European ancestry-matched control individuals selected from the large population-based cohort of UK Biobank (Pairo-Castineira *et al.*, 2021). A logistic regression model was used for each of the 4,380,209 SNPs individually tested for statistical significance.

We computed the density $\rho(r)$ of all the statistical test-associated $\chi^2_{\nu_1=1}$ r-statistics, and used a small 6-term Laguerre polynomial expansion for the entropic convex dual $\lambda(r)$. The resulting Gibbs prior was used to model the empirical NHST p-value empirical density $\rho(p(r))$ in terms of the Bayes Factor, that is,

$$BF(p(r)) = \int_\Lambda T^{\chi^2}_{\nu_1}(p(r)|\Lambda)\ \pi_L^{\left(-\frac{1}{2}\right)}(\Lambda)\ \mathrm{d}\Lambda. \tag{92}$$

As illustrated in Figure 9.1, Gibbs priors enable modeling of generic non-uniform empirical p-value densities via the Bayes Factor $BF(p)$. NHST statistical significance is replaced in the Bayesian framework by strength of Bayesian evidence as assessed by magnitude of $BF(p)$, which in turns allows computation of a local false discovery rate (Stephens, 2017)

$$\mathrm{fdr}(p) = 1/(1 + BF(p). \tag{93}$$

10. Conclusion

All four noncentral univariate t, normal, F, and χ^2 distributions $\rho(r|r_o)$ — with r_o the corresponding noncentrality parameter — can be constructed in a modular fashion by multiplying their central counterparts $\rho(r|r_o = 0)$ with a translating factor $T(r|r_o)$ effecting a central distribution translation, that is, we have that $\rho(r|r_o) = T(r|r_o)\ \rho(r|0)$. This modular construction applies to both the noncentral ultraspherical distributions $\upsilon(r|r_o)$ and their classical counterparts $\rho(r|r_o)$.

The classical noncentral distributions $\rho(r|r_o)$ all have a submodular decomposition for their translation factor $T(r|r_o) = E(r|r_o)\ e^{-r_0^2/2}$, with $E(r|r_o)$ a generalized exponential. See Appendix B for details. Unfortunately, such a factorization does not regroup terms of similar order in the noncentrality parameter r_o, thus does not allow expansion of the translation factor $T(r|r_o)$ in terms of orthogonal polynomials (Ismail *et al.*, 2005; Szegö, 1939).

An alternative factorization of the translation factor is possible if one considers the noncentral ultraspherical distributions. With the central distribution $\rho(r|0)$ playing the role of orthogonal polynomial family-defining weight function $w(r)$, the translation factor $T(r|r_o)$ for the noncentral ultraspherical t distribution was shown to be a generating function for the Gegenbauer polynomials, the translation factor for the noncentral normal distribution was argued to be the generating function for the Hermite polynomials, the translation factor for the noncentral ultraspherical F distribution was shown to have an expansion in terms of Jacobi polynomials, and the translation factor for the noncentral χ^2 distribution was shown to have an expansion in terms of Laguerre polynomials.

The parametric Bayesian approach to the inverse problem of determining the prior $\pi(r_o)$ underlying a generic empirical distribution $\rho(r) = \int \rho(r|r_o)\ \pi(r_o)\ \mathrm{d}r_o$ ultimately depends on the accurate and expedient determination of the entropic convex dual of $\rho(r)$ (Le Blanc, 2022). As demonstrated herein, expansion of the entropic convex dual $\lambda(r)$ in terms of members of an orthogonal polynomial family $\{P_n(r)\}_{n=0}^\infty$ defined by the central density $w(r) \propto \rho(r|r_o = 0)$, with expansion coefficients given by

$$\lambda_n = \frac{1}{\|P_n\|^2} \int_{\mathcal{R}} P_n(r)\ \lambda(r)\ w(r)\ \mathrm{d}r,$$

significantly reduced the computational burden of its determination by generally requiring computation of only a small subset of low-order coefficients.

We surveyed the literature concerned with the use of orthogonal polynomial families in optimization theory. Our search identified the average-case analysis of random quadratic problems by Pedregosa & Scieur (2020) who postulated specific parametric models for the expected spectrum of the Hessian matrix H eigenvalue distribution. Their work uses first order gradient methods, and draws on the polynomial-based iterative methods by Fischer (2011). The residual orthogonal polynomial families $\{P_n(r)\}$ are defined such that the model error expectation at the n^{th} iteration is proportional to the squared norm of the polynomials,

$$\|P_n\|^2 = \int_r P_n^2(r)\ w_s(r)\ \mathrm{d}r, \tag{94}$$

with $w_s(r)$ the postulated expected spectral distribution of the empirical spectral distribution of H. They considered three different parametric models for the expected spectral distribution $w_S(r)$, and identified for each model the corresponding residual orthogonal polynomial family. Since orthogonal polynomial families are defined by their weight function w, the Marchenko-Pastur spectral density lead to the shifted Chebyshev polynomials of the second kind (Gautschi & Milovanović, 2021), the exponential spectral density to the generalized Laguerre polynomials (Olver *et al.*, 2021), and the uniform spectral density to the shifted Legendre polynomials (Olver *et al.*, 2021), respectively. Parallels can be drawn with our work: modeling complexity resides in the choice of the model-defining density w, while optimization objectives are carried out in terms of the associated orthogonal polynomial family.

We conclude by drawing attention of the reader on the similarities between the modular construction for the noncentral ultraspherical t distribution and the noncentral normal distribution for which the multiplicative translation factors are generating functions for the Gegenbauer and Hermite orthogonal polynomial families, respectively, and the constructions of quantum coherent states in terms of similar generating functions (Ali & Ismail, 2012; Mojaveri & Dehghani, 2015). Recall that coherent states have been introduced first as quasi-classical states in quantum mechanics and quantum optics, but have ultimately engendered an important corpus of developments in Mathematical Physics (Wikipedia contributors, 2021). The present work should be considered part of this continuum.

Appendix

A. von Mises-Fisher-Langevin distribution & modified Bessel function of the first kind

When rewritten in terms of the central ultraspherical t-distribution (7), the modified Bessel function of the first kind I_μ is given by the integral representation (Segun & Abramowitz, 1965)

$$I_\mu(\kappa) = \frac{(\kappa/2)^\mu}{\Gamma(\mu+1)} \int_{\psi=0}^{\pi} e^{\kappa\cos\psi}\ \upsilon^t_{2\mu+1}(\psi|0)\ \mathrm{d}\psi. \tag{95}$$

In turn, the von Mises-Fisher-Langevin (vMFL) distribution (Hartman & Watson, 1974) can be defined as the normalized integrand of the latter:

$$\begin{aligned} \upsilon_\nu^{\mathrm{vMFL}}(\psi|\ \kappa) &= \frac{(\kappa/2)^{\frac{\nu-1}{2}}}{\Gamma(\frac{\nu+1}{2})\ I_{(\nu-1)/2}(\kappa)}\ e^{\kappa\cos\psi}\ \upsilon_\nu^t(\psi|0) && (96)\\ &= \frac{1}{\mathcal{I}_{(\nu-1)/2}(\kappa^2)}\ e^{\kappa\cos\psi}\ \upsilon_\nu^t(\psi|0), \qquad 0\le\psi\le\pi, \quad \kappa\ge 0, && (97)\end{aligned}$$

where we have made use of the definition of the normalized modified Bessel function of the first kind (András & Baricz, 2008)

$$\mathcal{I}_\mu(\kappa) = 2^\mu \kappa^{-\mu/2}\ \Gamma(\mu+1)\ I_\mu(\sqrt{\kappa}) = \int_{\psi=0}^{\pi} e^{\sqrt{\kappa}\cos\psi}\ \upsilon_{2\mu+1}^t(\psi|0)\ \mathrm{d}\psi. \tag{98}$$

The parameter κ is usually referred as the concentration parameter. When $\kappa = 0$, the vMFL distribution $\upsilon_\nu^{\mathrm{vMFL}}(\psi|0)$ simplifies to central hypersphere t distribution $\upsilon_\nu^t(\psi|0)$. As is done for the latter, it is understood that integration has been performed over the generalized azimuthal coordinates orthogonal to the polar axis, subspace unto which the vMFL density is uniform. We need a generalizable expression for the normalized modified Bessel function of the first kind

$$\mathcal{I}_{(\nu_1-2)/2}(\kappa) = \int_{\psi=0}^{\pi} e^{\sqrt{\kappa}\cos\psi}\ \upsilon_{\nu_1-1}^t(\psi)\ \mathrm{d}\psi, \quad \nu_1\ge 1. \tag{99}$$

The special case $\nu_1 = 1$ is given by the hyperbolic cosine of the square root of its argument, that is,

$$\mathcal{I}_{-1/2}(\kappa) = \cosh\sqrt{\kappa} = \sum_{j\ \text{even}\ \ge\ 0}^{\infty} \frac{\kappa^{j/2}}{j!}. \tag{100}$$

The case $\nu_1 = 2$

$$\mathcal{I}_o(\kappa) = \frac{1}{\pi}\int_{\psi=0}^{\pi} e^{\sqrt{\kappa}\cos\psi}\ \mathrm{d}\psi \tag{101}$$

does not have a closed expression in terms of analytic functions, but its Maclaurin expansion is given below in equation (103). The case $\nu_1 = 3$ is easily computed to be given by

$$\mathcal{I}_{1/2}(\kappa) = \frac{\sinh\sqrt{\kappa}}{\sqrt{\kappa}}. \tag{102}$$

One could use recursion formulæ to obtain expressions for the higher order functions $\mathcal{I}_{(\nu_1-2)/2}$, $\nu_1 > 2$, in terms of derivatives of $\mathcal{I}_{-1/2}$ ($\nu_1 = 1$) and $\mathcal{I}_o$ ($\nu_1 = 2$), but the algebra becomes rapidly prohibitive. See, for example, András & Baricz (2008). Instead, observe that the normalized modified Bessel function $\mathcal{I}_{(\nu_1-2)/2}(\kappa)$, $\nu_1\ge 2$, can be provided with a simple Maclaurin expansion by expanding the exponential term $e^{\sqrt{\kappa}\cos\psi}$ in the integral (99), and using definition of the ultraspherical central F-distribution (51) before carrying out the integral over ψ. One easily computes

$$\begin{aligned} \mathcal{I}_{(\nu_1-2)/2}(\kappa) &= \frac{\Gamma(\nu_1/2)}{\Gamma(1/2)} \sum_{j\ \text{even}\ \ge\ 0}^{\infty} \frac{\Gamma((j+1)/2)}{\Gamma((j+\nu_1)/2)}\ \frac{\kappa^{j/2}}{j!}, \quad \nu_1\ge 1, && (103)\\ &= \sum_{j=0}^{\infty} \frac{\Gamma(\nu_1/2)}{\Gamma(j+\nu_1/2)}\ \frac{(\kappa/4)^j}{j!} = {}_0F_1\Big(;\frac{\nu_1}{2};\frac{\kappa}{4}\Big), \end{aligned}$$

Noncentral	=	Translation T	$\times$	Central
$\upsilon^F_{(\nu_1,\ \nu_2)}(\theta\|\ \Lambda)$ (50)	=	$T^F_{(\nu_1,\ \nu_2)}(\theta\|\ \Lambda)$ (52)	$\times$	$\upsilon^F_{(\nu_1,\ \nu_2)}(\theta\|\ 0)$ (51)
$\downarrow \nu_2 \to \infty$				
$\rho^{\chi^2}_{\nu_1}(r\|\ \Lambda)$ (78)	=	$E^{\chi^2}_{\nu_1}(r\|\ \Lambda)\ e^{-\Lambda/2}$ (79)	$\times$	$\rho^{\chi^2}_{\nu_1}(r\|\ 0)$ (74)
$\uparrow \nu_2 \to \infty$				
$\rho^F_{(\nu_1,\ \nu_2)}(\theta\|\ \Lambda)$ (104)	=	$E^F_{(\nu_1,\ \nu_2)}(\theta\|\ \Lambda)\ e^{-\Lambda/2}$ (105)	$\times$	$\upsilon^F_{(\nu_1,\ \nu_2)}(\theta\|\ 0)$ (51)
$\upsilon^t_{\nu_2}(\theta\|\ \delta)$ (6)	=	$T^t_{\nu_2}(\theta\|\ \delta)$ (8)	$\times$	$\upsilon^t_{\nu_2}(\theta\|\ 0)$ (7)
$\downarrow \nu_2 \to \infty$				
$\rho^N(r\|\ \delta)$ (36)	=	$E^N(r\|\ \delta)\ e^{-\delta^2/2}$ (34)	$\times$	$\rho^N(r\|\ 0)$ (32)
$\uparrow \nu_2 \to \infty$				
$\rho^t_{\nu_2}(\theta\|\ \delta)$ (106)	=	$E^t_{\nu_2}(\theta\|\ \delta)\ e^{-\delta^2/2}$ (107)	$\times$	$\upsilon^t_{\nu_2}(\theta\|\ 0)$ (7)

Table B.1: **Modular expressions for the univariate noncentral distributions.** By convention, we designate the densities derived in terms of the translated hypersphere by the Greek letter υ, and the densities derived in terms of the normal and χ^2 distributions by the Greek letter ρ. Each noncentral univariate distribution repertoried in the left end column is given as the product of its corresponding central distribution in the right end column, times a translation factor T given in the middle column. Each modular component is provided with a reference to its equation in the manuscript. As illustrated by the vertical arrows, both noncentral t and F distributions, either in their ultraspherical υ or classical ρ versions, converge to the noncentral normal and χ^2 distributions, respectively. All the classical distributions have a submodular decomposition for their translation factor $T(r|r_o) = E(r|r_o)\ e^{-r_0^2/2}$, with $E(r|r_o)$ a generalized exponential, not suitable for expansion in terms of orthogonal polynomial families.

which encompasses the special case $\mathcal{I}_{-1/2}$ $(\nu_1 = 1)$. The normalized modified Bessel function $\mathcal{I}_{(\nu_1-2)/2}(\kappa)$ behaves as a generalized exponential function. They appear in the classical expression for the noncentral χ^2 distributions.

B. Modular expressions for the classical noncentral distributions

The normalized modified Bessel function of the first kind was provided in (103) with a novel Maclaurin expansion. We shall see in this Appendix that this expression generalizes to all the noncentral distributions when derived in terms of the normal and the χ^2 distributions. The noncentral F distribution has been classically defined as the ratio of a variable transforming according to a noncentral χ^2 distribution with ν_1 degrees of freedom and noncentrality parameter Λ, over that of a variable transforming according to a central χ^2 distribution with ν_2 degrees of freedom. See, for example, Walck (1996). Using the present formalism, one realizes that the resulting noncentral F distribution expansion can be rewritten as the product

$$\rho^F_{(\nu_1,\nu_2)}(\theta|\Lambda) = E^F_{(\nu_1,\nu_2)}(\theta|\Lambda)\ e^{-\Lambda/2}\ \upsilon^F_{(\nu_1,\nu_2)}(\theta|0) \tag{104}$$

where

$$
\begin{aligned}
E^F_{(\nu_1,\nu_2)}(\theta|\Lambda) &= \frac{\Gamma(\nu_1/2)}{\Gamma(1/2)} \sum_{j \text{ even } \geq 0}^{\infty} \frac{\Gamma((j+1)/2)}{\Gamma((j+\nu_1)/2)} \frac{\Gamma((j+\nu_1+\nu_2)/2)}{\Gamma((\nu_1+\nu_2)/2)} \frac{(\sqrt{2\Lambda}\cos\theta)^j}{j!} \\
&= \sum_{j=0}^{\infty} \frac{\Gamma(\nu_1/2)}{\Gamma(j+\nu_1/2)} \frac{\Gamma(j+(\nu_1+\nu_2)/2)}{\Gamma((\nu_1+\nu_2)/2)} \frac{(\Lambda\cos^2\theta/2)^j}{j!} \qquad (105) \\
&= {}_1F_1\left(\frac{\nu_1+\nu_2}{2}; \frac{\nu_1}{2}; \frac{\Lambda\cos^2\theta}{2}\right).
\end{aligned}
$$

Performing the two changes of variables $\theta = \frac{\pi}{2} - \phi$ followed by $r = \nu_2\phi^2$ used in section 8 to derive the noncentral χ^2 distribution from the noncentral F distribution, we have that

$$
\lim_{\nu_2\to\infty} E^F_{(\nu_1,\nu_2)}(r|\Lambda) = E^{\chi^2}_{\nu_1}(r|\Lambda) = \mathcal{I}_{(\nu_1-2)/2}(\Lambda r) = {}_0F_1\left(; \frac{\nu_1}{2}; \frac{\Lambda r}{4}\right),
$$

that is, we retrieve the noncentral χ^2 distribution (78) as the limiting distribution of the noncentral F distribution (104), as should be.

Similarly, the noncentral t distribution has been classically defined as the ratio of a variable transforming according to a noncentral normal distribution with noncentrality parameter δ, over that of a variable transforming according to a central χ distribution with ν_2 degrees of freedom. See again Walck (1996). Using the present formalism, one can similarly conclude that the resulting noncentral t distribution expansion can be rewritten as the product

$$
\rho^t_{\nu_2}(\theta|\delta) = E^t_{\nu_2}(\theta|\delta)\ e^{-\delta^2/2}\ \upsilon^t_{\nu_2}(\theta|0), \qquad (106)
$$

where

$$
\begin{aligned}
E^t_{\nu_2}(\theta|\delta) &= \sum_{j=0}^{\infty} \frac{\Gamma((j+1+\nu_2)/2)}{\Gamma((1+\nu_2)/2)} \frac{(\sqrt{2}\delta\cos\theta)^j}{j!} \qquad (107) \\
&= {}_1F_1\left(\frac{\nu_2+1}{2}; \frac{1}{2}; \frac{\delta^2\cos^2\theta}{2}\right) \\
&\qquad + \sqrt{2}\ \delta\cos\theta\ \frac{\Gamma(\frac{\nu_2+2}{2})}{\Gamma(\frac{\nu_2+1}{2})}\ {}_1F_1\left(\frac{\nu_2+2}{2}; \frac{3}{2}; \frac{\delta^2\cos^2\theta}{2}\right),
\end{aligned}
$$

which equates equation (105) above when $\nu_1 = 1$ except for its summation involving both even and odd integers. Note that j is divided by 2 in (107), explaining the need in this case for the sum over two hypergeometric functions. Performing the two changes of variables $\theta = \frac{\pi}{2} - \phi$ followed by $r = \sqrt{\nu_2}\phi$ used in section 4 to derived the noncentral normal distribution from the noncentral t distribution, we have that

$$
\begin{aligned}
\lim_{\nu_2\to\infty} E^t_{\nu_2}(r|\delta) &= {}_0F_1\left(; \frac{1}{2}; \frac{\delta^2 r^2}{4}\right) + \delta r\ {}_0F_1\left(; \frac{3}{2}; \frac{\delta^2 r^2}{4}\right) \\
&= \cosh\delta r + \sinh\delta r = e^{\delta r} = E^N(r|\delta)
\end{aligned}
$$

that is, we retrieve the noncentral normal distribution (36) as the limiting distribution of the noncentral t distribution (106), as should be.

To summarize, all four univariate noncentral t, normal, F, and χ^2 distributions as classically derived in terms of the normal and χ^2 distributions have been provided with similar modular expressions, using a unified formalism in which the Maclaurin expansions for their respective generalized exponential factors E are successively simpler versions of the Maclaurin expansion for the noncentral F distribution generalized exponential factor $E^F_{(\nu_1,\nu_2)}$: we have that

$$E^{\chi^2}_{\nu_1} \underset{\nu_2\to\infty}{\longleftarrow} E^F_{(\nu_1,\nu_2)} \underset{\nu_1=1}{\longrightarrow} E^t_{\nu_2} \underset{\nu_2\to\infty}{\longrightarrow} E^N, \tag{108}$$

with

$$\begin{aligned}
E^F_{(\nu_1,\nu_2)}(\theta|\Lambda) &= \frac{\Gamma(\nu_1/2)}{\Gamma(1/2)} \sum_{j \text{ even } \geq 0}^{\infty} \frac{\Gamma((j+1)/2)}{\Gamma((j+\nu_1)/2)} \frac{\Gamma((j+\nu_1+\nu_2)/2)}{\Gamma((\nu_1+\nu_2)/2)} \frac{(\sqrt{2\Lambda}\cos\theta)^j}{j!}, \\
E^{\chi^2}_{\nu_1}(r|\Lambda) &= \frac{\Gamma(\nu_1/2)}{\Gamma(1/2)} \sum_{j \text{ even } \geq 0}^{\infty} \frac{\Gamma((j+1)/2)}{\Gamma((j+\nu_1)/2)} \frac{(\Lambda r)^{j/2}}{j!} = \mathcal{I}_{(\nu_1-2)/2}(\Lambda r), \\
E^t_{\nu_2}(\theta|\delta) &= \sum_{j=0}^{\infty} \frac{\Gamma((j+1+\nu_2)/2)}{\Gamma((1+\nu_2)/2)} \frac{(\sqrt{2}\delta\cos\theta)^j}{j!}, \\
E^N(r|\delta) &= \sum_{j=0}^{\infty} \frac{(\delta r)^j}{j!} = e^{\delta r},
\end{aligned}$$

and with the understanding that the respective Maclaurin summations involve either only the even integers for the F and χ^2 cases, or both odd and even integers for the t and normal cases. All of these analytical expansions have been verified to numerically reproduce the corresponding noncentral distributions as implemented in MATLAB®.

C. Hyperspherical and classical t and F distributions' Kullback-Leibler divergence

We shall consider in this Appendix the Kullback-Leibler divergence between the noncentral ultraspherical and classical t and F distributions, together with their conditions of applicability. From the preceding sections and as indicated by the vertical arrows in Table B.1, we already know that, in the limit of high degree of freedom ν_2, both the noncentral ultraspherical t distribution $\upsilon^t_{\nu_2}(\theta|\delta)$ and its classical counterpart $\rho^t_{\nu_2}(\theta|\delta)$ converge unto the noncentral normal distribution, while the noncentral ultraspherical F distribution $\upsilon^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ and its classical counterpart $\rho^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ both converge unto the noncentral χ^2 distribution. Consequently, the Kullback-Leibler divergence between the ultraspherical and classical noncentral t and F distributions is expected to vanish when $\nu_2 \to \infty$.

Consider then the central t and F distributions, that is, the distributions with noncentrality parameter δ and Λ set to zero, respectively. By definition and as

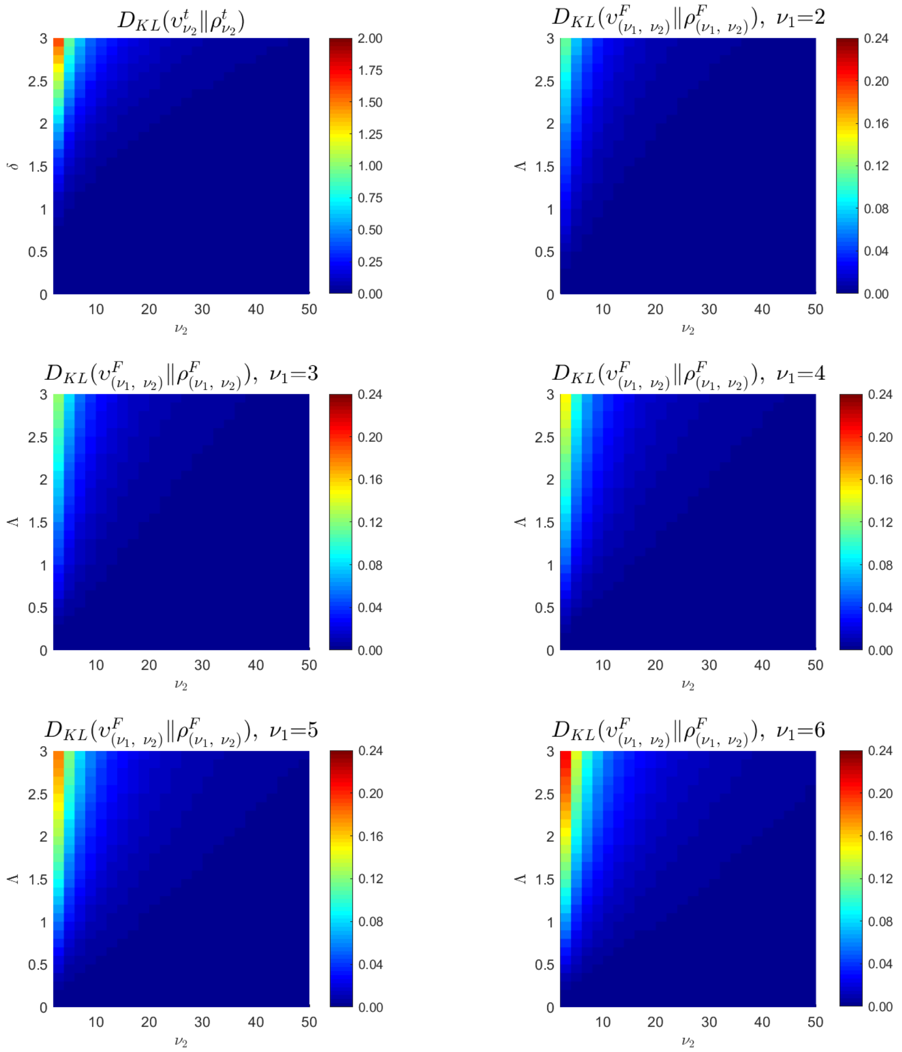

Figure C.1: **Symmetric nonnegative Kullback-Leibler divergence** between the noncentral ultraspherical $\upsilon^t_{\nu_2}(\theta|\delta)$ and classical $\rho^t_{\nu_2}(\theta|\delta)$ $t-$distributions in the left upper corner panel, and between the noncentral ultraspherical $\upsilon^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ and classical $\rho^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ F distributions for the listed parameter ν_1 in all the other panels.

argued in the introduction, the central ultraspherical t (7) and F (51) distributions are simply projections of a central hypersphere uniform density on a polar axis and a secant hyperplane, respectively. Now, in an ANOVA, the t and F statistics are defined as the ratio of the between-class variance to the within-class variance of an observation vector, and this ratio is independent of the observation vector's length. Computation of the statistics will therefore project any observation vector unto a cotangent vector to a referential central hypersphere, as can be deduced from the changes of variable (5) and (49). It follows that vectors sampled from a central normal distribution $N(0, I_n)$ in $\mathbb{R}^n$ will, by rotational invariance, project unto a uniform density on the referential central hypersphere $\mathcal{S}^{n-1}$ (Vershynin, 2018, p. 53). Consequently, the central ultraspherical and classical t and F distributions will share the same central densities $\upsilon^t_{\nu_2}(\theta|0)$ and $\upsilon^F_{(\nu_1,\nu_2)}(\theta|0)$, respectively. See their respective entries in Table B.1. We thus have the added knowledge that the Kullback-Leibler divergence between the ultraspherical and classical t and F distributions vanishes when their respective noncentrality parameter δ and Λ is set to zero.

Consider now generic noncentral distributions for which the noncentrality parameters δ or Λ are non-vanishing. We have seen in sections 4 and 8 that the noncentral t and F ultraspherical distributions (6) and (50) are, by construction, projections of a uniform but translated ultraspherical distribution on the referential central hypersphere. In counterpart, we saw in Appendix B that the noncentral t and F distributions (106) and (104) have been classically derived in terms of the marginal of the joint distribution of a noncentral normal distribution times that of a central χ_{ν_2} distribution for the former, and the marginal of the joint distribution of a noncentral $\chi^2_{\nu_1}$ distribution times that of a central $\chi^2_{\nu_2}$ distribution for the later. This implies that vectorial samples are drawn from a translated normal distribution in $\mathbb{R}^n$, with $n = \nu_2 + 2$ for the former, and $n = \nu_1 + \nu_2 + 1$ for the later. As a result, and since rotational invariance cannot be invoked for translated distributions, the noncentral ultraspherical and the noncentral classical t and F distributions are expected to diverge from one another since the vectorial observations are drawn from different distributions which will project differently on the referential central hypersphere. Of note, this divergence arises only through their translation factors T as listed in Table B.1. The fundamental issue is thus of deciding the conditions of applicability of the noncentral ultraspherical distributions versus those of the classically derived noncentral distributions. Stated otherwise, when should one consider the vectorial observations as being drawn from either a uniform distribution lying on a translated hypersphere $\mathcal{S}^{n-1}$, or from a normal distribution translated in $\mathbb{R}^n$?

It turns out that the latter distinction is irrelevant in high dimensional spaces when one takes into account the counterintuitive properties of the hypersphere therein. Indeed, since the density in high dimension n of a normal distribution $N(0, I_n)$ in $\mathbb{R}^n$ concentrates not as a ball centered on the origin but rather within a thin spherical shell of thickness $O(1)$ around the hypersphere of radius $\sqrt{n}$ (Vershynin, 2018, p. 53), it is expected that the nonnegative symmetric Kullback–Leibler divergence

$$D_{KL}(\upsilon^t_{\nu_2}(\cdot|\delta)\|\rho^t_{\nu_2}(\cdot|\delta)) = \int_{\theta=0}^{\pi} [\upsilon^t_{\nu_2}(\theta|\delta) - \rho^t_{\nu_2}(\theta|\delta)] \times \ln \frac{\upsilon^t_{\nu_2}(\theta|\delta)}{\rho^t_{\nu_2}(\theta|\delta)} \, \mathrm{d}\theta \qquad (109)$$

between the noncentral ultraspherical and classical t distributions will be negligible in high dimension, with similar results for the divergence D_{KL} between the ultraspherical and classical noncentral F distributions. As can be seen in Figure C.1, notable Kullback–Leibler divergence for the noncentral t-distributions arise only for very low sample size $n = \nu_2 + 2$ in conjunction with large normalized effect size $\delta \gtrsim 2$. Since large normalized effect size are unusual in real life situations, and since one should strive to consider samples of size n large enough to compute with some accuracy both the between-class and within-class variances of an observation vector, Figure C.1 ascertains that the discrepancy between the ultraspherical $\upsilon^t_{\nu_2}(\theta|\delta)$ and the classical $\rho^t_{\nu_2}(\theta|\delta)$ noncentral $t-$distributions can be neglected in most circumstances. Similar conclusions apply to the discrepancy between the ultraspherical $\upsilon^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ and the classical $\rho^F_{(\nu_1,\nu_2)}(\theta|\Lambda)$ noncentral $F-$distributions (50) and (104). As a rule of thumb, in an ANOVA with number of class $c = \nu_1 + 1$ and number of samples $n = \nu_1 + \nu_2 + 1$, Figure C.1 indicates that a number of samples $n \sim (3 - 5) \times c$ should be sufficient to insure convergence of the noncentral ultraspherical and classical distributions for small effect size δ or Λ, with the noncentral t and F ultraspherical distributions (6) and (50) then considered to be good approximations to the noncentral classical t and F distributions (106) and (104). If, on the contrary, large effect size and low dimensions are being considered, one could readily specify the assumed underlying distribution and corresponding sample space for the problem at hand.

To summarize, the ultraspherical and the classical noncentral F and t distributions: correspond to projections of translated hypersphere and translated normal distributions, respectively; are identical in their central distribution versions when their noncentrality parameters are zero; converge in high dimensional spaces; but diverge in low dimension spaces and for large noncentrality parameters. These properties ultimately stem from the counterintuitive properties of the solid hypersphere which concentrates its volume on a thin ultraspherical shell in high dimensional spaces, allowing one to use the simpler compact ultraspherical distributions as surrogates for the classical noncentral t and F distributions in high dimensional spaces.

References

Ali, S Twareque, & Ismail, Mourad EH. 2012. Some orthogonal polynomials arising from coherent states. *Journal of Physics A: Mathematical and Theoretical*, **45**(12), 125203.

András, Szilárd, & Baricz, Árpád. 2008. Properties of the probability density function of the non-central chi-squared distribution. *Journal of mathematical analysis and applications*, **346**(2), 395–402.

Fischer, Bernd. 2011. *Polynomial based iteration methods for symmetric linear systems*. SIAM.

Gautschi, Walter, & Milovanović, Gradimir V. 2021. Orthogonal polynomials relative to a generalized Marchenko–Pastur probability measure. *Numerical Algorithms*, 1–17.

Gorroochurn, Prakash. 2016. *Classic Topics on the History of Modern Mathematical Statistics*. Wiley Online Library.

Hartman, Philip, & Watson, Geoffrey S. 1974. Normal distribution functions on spheres and the modified Bessel functions. *The Annals of Probability*, 593–607.

Ismail, Mourad, Ismail, Mourad EH, & van Assche, Walter. 2005. *Classical and quantum orthogonal polynomials in one variable*. Vol. 13. Cambridge university press.

Le Blanc, Richard. 2019. Bayesian Analysis on a Noncentral Fisher–Student's Hypersphere. *The American Statistician*, **73**(2), 126–140.

Le Blanc, Richard. 2022. Entropic convex duality in the determination of data-constrained kernel-based Bayes-Jaynes priors. *Journal of Convex Analysis*.

Mojaveri, B, & Dehghani, A. 2015. New generalized coherent states arising from generating functions: A novel approach. *Reports on Mathematical Physics*, **75**(1), 47–61.

Olver, F. W. J., Olde Daalhuis, A. B., Lozier, D. W., Schneider, B. I., Boisvert, R. F., Clark, C. W., Miller, B. R., Saunders, B. V., Cohl, H. S., & McClain, M. A. 2021. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.1.3 of 2021-09-15.

Pairo-Castineira, Erola, Clohisey, Sara, Klaric, Lucija, Bretherick, Andrew D, Rawlik, Konrad, Pasko, Dorota, Walker, Susan, Parkinson, Nick, Fourman, Max Head, Russell, Clark D, *et al.* 2021. Genetic mechanisms of critical illness in Covid-19. *Nature*, **591**(7848), 92–98.

Pedregosa, Fabian, & Scieur, Damien. 2020. Acceleration through spectral density estimation. *Pages 7553–7562 of: International Conference on Machine Learning*. PMLR.

Segun, A, & Abramowitz, M. 1965. Handbook of mathematical functions. *Appl. Math. Ser*, **55**.

Shizgal, Bernie D, & Jung, Jae-Hun. 2003. Towards the resolution of the Gibbs phenomena. *Journal of Computational and Applied Mathematics*, **161**(1), 41–65.

Stephens, Matthew. 2017. False discovery rates: a new deal. *Biostatistics*, **18**(2), 275–294.

Szegö, Gabor. 1939. *Orthogonal polynomials*. Vol. 23. American Mathematical Soc.

Tiku, ML. 1965. Laguerre series forms of non-central χ^2 and F distributions. *Biometrika*, **52**(3/4), 415–427.

Van Aubel, Andrea, & Gawronski, Wolfgang. 2003. Analytic properties of noncentral distributions. *Applied mathematics and computation*, **141**(1), 3–12.

Vershynin, Roman. 2018. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Vol. 47. Cambridge University Press.

Walck, Christian. 1996. *Hand-book on statistical distributions for experimentalists*. Tech. rept.

Wikipedia contributors. 2021. *Coherent states in mathematical physics — Wikipedia, The Free Encyclopedia*. [Online; accessed 4-January-2022].

Yu, Yaming. 2011. The Shape of the Noncentral Chi-square Density. *arXiv preprint arXiv:1106.5241*.