



# Article Jaynes-Gibbs Entropic Convex Duals and Orthogonal Polynomials

Richard Le Blanc 匝

Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke 1, 3001, 12 ème Avenue Nord, Sherbrooke, QC J1H 5N4, Canada; richard.le.blanc@usherbrooke.ca

Abstract: The univariate noncentral distributions can be derived by multiplying their central distributions with translation factors. When constructed in terms of translated uniform distributions on unit radius hyperspheres, these translation factors become generating functions for classical families of orthogonal polynomials. The ultraspherical noncentral *t*, normal *N*, *F*, and  $\chi^2$  distributions are thus found to be associated with the Gegenbauer, Hermite, Jacobi, and Laguerre polynomial families, respectively, with the corresponding central distributions standing for the polynomial family-defining weights. Obtained through an unconstrained minimization of the Gibbs potential, Jaynes' maximal entropy priors are formally expressed in terms of the empirical densities' entropic convex duals. Expanding these duals on orthogonal polynomial bases allows for the expedient determination of the Jaynes–Gibbs priors. Invoking the moment problem and the duality principle, modelization can be reduced to the direct determination of the prior moments in parametric space in terms of the Bayes factor's orthogonal polynomial expansion coefficients in random variable space. Genomics and geophysics examples are provided.

**Keywords:** noncentral distributions; orthogonal polynomials; Bayesian inference; Jaynes' maximal entropy principle; Gibbs prior; entropic convex dual



Citation: Le Blanc, R. Jaynes-Gibbs Entropic Convex Duals and Orthogonal Polynomials. *Entropy* 2022, 24, 709. https://doi.org/ 10.3390/e24050709

Academic Editor: Nikolai Leonenko

Received: 8 March 2022 Accepted: 10 May 2022 Published: 16 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

# 1. Introduction

We shall argue that the four noncentral univariates *t*, *F*, the normal *N*, and  $\chi^2$  distributions  $\rho(r|r_o)$ , with *r* being the one-dimensional random space variable, and  $r_o$  the one-dimensional 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,

$$\rho(r|r_o) = T(r|r_o) \rho(r|0).$$

In statistical parlance, the noncentral distributions are needed to estimate or modelize effect sizes [1]. With the exception of the normal distribution, these translations are non-shape-preserving. The derivation of the translation factors  $T(r|r_0)$  can be carried out in two manners, depending on whether primacy is put upon translated uniform density distributions on unit radius hyperspheres, as is done in this manuscript, or on translated normal distributions, as done classically. We shall review both derivations herein, with emphasis on the hyperspherical distributions.

We choose to place primacy upon the simple uniform density distribution on the unit radius hypersphere  $S^{\nu}$ , where  $\nu$  stands for either the dimension of the hypersphere surface  $S^{\nu} \in \Re^{\nu+1}$ , or, equivalently, the degrees of freedom (dof) of the specified distribution. It is known that the projection of a unit radius uniform density hyperspherical distribution on  $S^{\nu}$  on any given polar axis readily provides us with the central *t* distribution [2], and that such a projection converges with the central normal distribution N(0, 1) with null mean when  $\nu$  tends to infinity [3] (p. 59). This observation provides us with the needed building principle used throughout this manuscript: use the uniform density distribution on the unit radius hypersphere  $S^{\nu} \in \Re^{\nu+1}$  to derive modular expressions for the central and noncentral t, F, N, and  $\chi^2$  distributions. In order to proceed, one needs to master some very simple notions concerning the hypersphere geometry. The projection of a random unit vector **x** on the unit radius hypersphere  $S^{\nu_2}$  on any given unit polar axis **p** naturally defines a polar angle  $\theta$  through the scalar product  $\cos \theta = \mathbf{x} \cdot \mathbf{p}$ . As such, the central *t* distribution with  $\nu_2$  degrees of freedom can be drawn on the compact support  $-1 < \cos \theta < 1$ , on which it acquires a simple expression in trigonometric terms: it is simply proportional to  $\sin^{\nu_2-1}\theta$  after integration of the azimuthal coordinates. See Saville and Wood [2] for a very extensive digression on the subject. On the familiar sphere  $S^2 \in \Re^3$ , the latter provides us with the well-known spherical surface element  $\sin \theta \, d\theta$  after the integration of the azimuthal coordinate. Similarly, the central  $F^{(v_1, v_2)}$  distribution becomes proportional to  $\cos^{\nu_1 - 1} \theta \sin^{\nu_2 - 1} \theta$  on the compact domain  $0 \le \cos \theta < 1$ , where  $\theta$  is the angle between a random vector **x** on the hypersphere  $S^{\nu_1+\nu_2}$  and a secant hyperplane defining the subspace  $S^{\nu_1}$ . In statistical parlance, the  $S^{\nu_1}$  and  $S^{\nu_2}$  subspaces refer to the between-class and within-class variance spaces in an analysis variance (ANOVA), as reflected by the F-statistic redefinition in trigonometric terms, found in Section 2. As it turns out, these two central uniform hyperspherical distributions are all that is needed to proceed with the derivation of all the univariate noncentral t, F, N, and  $\chi^2$  distributions where, for uniformity of designation, a normal distribution  $N(\delta, 1)$  with non-vanishing mean  $\delta$  and unit variance will be simply referred to as the noncentral normal distribution. Finally, in the theory of orthogonal polynomials, the designation ultraspherical polynomials—also known as Gegenbauer polynomials [4]—has prevailed over that of hyperspherical polynomials, and we shall abide by this nomenclature. In order to distinguish the ultraspherical t and Fdistributions from the ones derived classically from the normal N and  $\chi^2$  distributions, we shall designate the former densities by the Greek letter v (upsilon), as in UTEPGPaúpa (ypersfaira or hypersphere in English), and the latter densities by the Greek letter  $\rho$ .

In Bayesian inference, the determination of the Bayes prior is referred to as an inverse problem, and Jaynes' data-constrained maximal entropy priors provide a principled solution to this inverse problem [5–7]. The maximal entropy is reached by minimizing the Gibbs potential, and the solution to this optimization problem requires the determination of the empirical density entropic convex dual. See Le Blanc [8] for an extensive review on the subject, with references. As derived from translated normal distributions, the classical noncentral distributions all have a submodular decomposition for their translation factor of the form

$$T(r|r_0) = E(r|r_0) e^{-r_0^2/2},$$

with  $E(r|r_o)$  being a generalized hypergeometric function, a decomposition which does not readily provide regrouping of terms of similar order in the noncentrality parameter  $r_o$ . Conversely, the translation factors for the noncentral t, the normal N, F, and  $\chi^2$  distributions, as derived from translated uniform distributions on unit radius hyperspheres, are found to be generating functions for the Gegenbauer, Hermite, Laguerre, and Jacobi orthogonal polynomial families  $\{P_n\}_{n=0}^{\infty}$ , respectively, and intrinsic properties of the orthogonal polynomials [9] allow for regrouping all terms of similar order in the noncentrality parameter  $r_o$ , that is,

$$\Gamma(r|r_o) = \sum_{n=0}^{\infty} c_n P_n(r) r_o^n,$$

with the constants  $c_n$  provided hereafter. To the best of our knowledge, the derivation of these translation factors and their identification as orthogonal polynomial family-generating functions has not been carried out before. As a consequence, one can expand the entropic convex duals on a generally small number of low-order orthogonal polynomials, an approach which greatly curtails the computational cost of determining the duals and, thus, the Jaynes–Gibbs priors. To the best of our knowledge, the expansion and discretization of the convex duals over orthogonal polynomial bases has not been proposed before. We adopt in this manuscript the convention that the polynomial family-defining weight functions

should be provided by the corresponding normalized central distributions, a convention which results in much simplified expressions for the norm of the orthogonal polynomials.

In parametric Bayesian modeling, the translation factor  $T(r|r_o)$  can be weighted by a Bayesian prior  $\pi(r_o)$  to obtain the Bayes factor

$$BF(r) = \int T(r|r_o) \ \pi(r_o) \ dr_o$$

for the generic superposition density

$$\rho(r) = \int \rho(r|r_o) \ \pi(r_o) \ dr_o = \int T(r|r_o) \ \rho(r|0) \ \pi(r_o) \ dr_o = BF(r) \ \rho(r|0).$$

We will, in this manuscript, identify the normalized central distribution  $\rho(r|0)$  with the orthogonal polynomial family-defining weight  $w(r) \equiv \rho(r|0)$  [9], with, as a result, the rewriting of any generic density  $\rho(r)$  as

$$\rho(r) = BF(r) w(r)$$

that is, the Bayes factor BF(r) can stand as a substitute to the generic density  $\rho(r)$ . With  $p(r) = \int^r \rho(r'|0) dr'$  standing for the cumulative density function of the respective central distributions  $\rho(r|0)$ , or, equivalently, the *p*-value of the null hypothesis statistical testing (NHST) procedure, we have that  $\rho(p) = BF(p)$ . That is, the Bayes factor BF(p) stands for the generally nonuniform *p*-value distribution  $\rho(p)$  of the above generic superposition density (see Le Blanc [10] for details of a proof), to be contrasted with the NHST framework which only considers the central distribution with its uninformative uniform *p*-value distribution. Now, if one's goal is only to model the density BF(p) or to compute the associated local false discovery rate fdr(p) = 1/(1 + BF(p)) [11], one then only needs prior moments to carry on with the modelization. Invoking the moment problem [12] and the duality principle [13], prior moments in parametric space will be shown to be readily provided by Bayes factor BF(r) orthogonal polynomial expansion coefficients in random variable space—that is,

$$\int_{\mathcal{R}_o} \pi(r_o) r_o^n \, \mathrm{d}r_o = \frac{1}{c_n \|P_n\|^2} \int_{\mathcal{R}} P_n(r') \, BF(r') \, w(r') \, \mathrm{d}r'.$$

The paper is organized as follows. In Section 2, we review, for the sake of completeness, the derivation of the classical univariate noncentral distributions as derived from translated normal distributions. In Section 3, the univariate ultraspherical noncentral *t*, the normal *N*, *F*, and  $\chi^2$  distributions are derived from translated uniform density distributions on unit radius hyperspheres, and are shown to be expressible as products of their central distribution times specific generating functions for the Gegenbauer, Hermite, Jacobi, and Laguerre orthogonal polynomial families, respectively. We argue, in Section 4, that the determination of the Gibbs priors in terms of empirical densities' entropic convex duals is much simplified when these duals are expanded on a small number of low-order orthogonal polynomials. We also discuss how prior moments in parametric space are directly provided by the Bayes factor orthogonal polynomial expansion coefficients in random variable space. Sections 5 and 6 are devoted to applications in genomics and geophysics, respectively.

#### 2. The Classical Noncentral Distributions

The four central *F*,  $\chi^2$ , *t*, and normal *N* distributions can be given by

$$v_F^{(\nu_1, \nu_2)}(\theta|0) = 2 \frac{\Gamma(\frac{\nu_1 + \nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \cos^{\nu_1 - 1}\theta \sin^{\nu_2 - 1}\theta, \quad 0 \le \cos\theta < 1,$$
(1a)

$$\rho_{\chi^2}^{(\nu_1)}(r|0) = \frac{1}{2^{\nu_1/2}\Gamma(\frac{\nu_1}{2})} r^{(\nu_1-2)/2} e^{-r/2}, \quad 0 \le r < \infty, \tag{1b}$$

$$\rho_t^{(\nu_2)}(\theta|0) = \frac{\Gamma(\frac{\nu_2+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{\nu_2}{2})} \sin^{\nu_2-1}\theta, \quad -1 < \cos\theta < 1, \tag{1c}$$

$$\rho_N(r|0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2}, \quad -\infty < r < \infty,$$
(1d)

respectively, where we have redefined the conventional t and F statistics so that

$$t = \sqrt{\nu_2} \frac{\cos \theta}{\sin \theta}, \qquad -1 < \cos \theta < 1,$$
 (2)

and

$$F = \frac{\nu_2}{\nu_1} \frac{\cos^2 \theta}{\sin^2 \theta}, \qquad 0 \le \cos \theta < 1, \tag{3}$$

so as to re-scale the corresponding t and F distributions on the specified finite compact domains. As discussed in the Introduction, the above central t and F distributions are projections of uniform distributions on the unit radius  $S^{\nu_2}$  and  $S^{\nu_1+\nu_2}$  hyperspheres, respectively. The reasons for such rescaling will become apparent when we link in Section 3 the ultraspherical noncentral t and F distributions with the Gegenbauer and Jacobi orthogonal polynomial families, respectively, families which both have similar finite compact support domains. It is straightforward to verify that the above central t and F distributions converge with the central normal N and  $\chi^2$  distributions, respectively, in the limit  $\nu_2 \rightarrow \infty$ : see the discussions leading to Equations (19) and (36) below. In geometric terms, the limit  $v_2 \to \infty$  allows one to restrict the study of the central *t* and *F* distributions on a restricted domain around the angle  $\theta = \pi/2$ , around which these distributions will concentrate their respective distribution weights in concordance with the notion that a high-dimensional hypersphere concentrates its surface on a narrow equatorial band at its equator [3]. In this limit, with the respective variable substitutions  $r = \sqrt{\nu_2 - 1} \cos \theta$  and  $r = (v_2 - 2) \cos^2 \theta$ , the normal N and  $\chi^2$  respective central distributions' definition domains consequently extend to infinity as  $\cos \theta \rightarrow 1$ , as reflected by the respective distribution domains provided above.

The noncentral *t* distribution has been classically derived by considering the ratio of a random variable which distributes according to a noncentral normal distribution  $N(\delta, 1)$  with non-vanishing mean  $\delta$ , over that of a random variable distributed according to a central  $\chi_{v_2}$  distribution with  $v_2$  degrees of freedom. Similarly, the noncentral *F* distribution is classically derived by considering the ratio of a noncentral  $\chi^2$  random variable with noncentrality parameter  $\Lambda$  and  $v_1$  degrees of freedom over that of a central  $\chi^2$  random variable with noncentrality parameter  $\Lambda$  and  $v_1$  degrees of freedom over that of a central  $\chi^2$  random variable with  $v_2$  degrees of freedom. See, for example, Walck et al. [14] for explicit steps for such derivations. As a result, as discussed in the Introduction, the classical distributions 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  $E(r|r_o)$  being a generalized hypergeometric function [15]. Upon recalling that the *F* and  $\chi^2$  noncentral parameter  $\Lambda$  stands for the square of the noncentral parameter  $\delta$  for the noncentral *t* and normal distributions, we have that

$$\rho_F^{(\nu_1, \nu_2)}(\theta|\Lambda) = E_F^{(\nu_1, \nu_2)}(\theta|\Lambda) \ e^{-\Lambda/2} \ \rho_F^{(\nu_1, \nu_2)}(\theta|0), \quad 0 \le \cos\theta < 1, \ 0 \le \Lambda < \infty,$$
(4a)

$$\rho_{\chi^2}^{(\nu_1)}(r|\Lambda) = E_{\chi^2}^{(\nu_1)}(r|\Lambda) \ e^{-\Lambda/2} \ \rho_{\chi^2}^{(\nu_1)}(r|0), \quad 0 \le r < \infty, \ 0 \le \Lambda < \infty,$$
(4b)

$$\rho_t^{(\nu_2)}(\theta|\delta) = E_t^{(\nu_2)}(\theta|\delta) \ e^{-\delta^2/2} \ \rho_t^{(\nu_2)}(\theta|0), \quad -1 < \cos\theta < 1, \ -\delta < \delta < \infty, \tag{4c}$$

$$\rho_N(r|\delta) = E_N(r|\delta) e^{-\delta^2/2} \rho_N(r|0), \quad -\infty < r < \infty, \ -\delta < \delta < \infty, \tag{4d}$$

with the respective factors'  $E(r|r_0)$  McLaurin expansions — computed following the steps elaborated by Walck et al. [14]—given by

$$E_F^{(\nu_1,\nu_2)}(\theta|\Lambda) = \frac{\Gamma(\nu_1/2)}{\Gamma(1/2)} \sum_{j \text{ even } \ge 0}^{\infty} \frac{\Gamma((j+1)/2)}{\Gamma((j+\nu_1)/2)} \frac{\Gamma((j+\nu_1+\nu_2)/2)}{\Gamma((\nu_1+\nu_2)/2)} \frac{(\sqrt{2\Lambda}\cos\theta)^j}{j!}$$
(5a)

$$= {}_{1}F_{1}\left(\frac{\nu_{1}+\nu_{2}}{2};\frac{\nu_{1}}{2};\frac{\Lambda\cos^{2}\theta}{2}\right),$$
(5b)

$$E_{\chi^2}^{(\nu_1)}(r|\Lambda) = \frac{\Gamma(\nu_1/2)}{\Gamma(1/2)} \sum_{j \text{ even } \ge 0}^{\infty} \frac{\Gamma((j+1)/2)}{\Gamma((j+\nu_1)/2)} \frac{(\Lambda r)^{j/2}}{j!} = \mathcal{I}^{(\nu_1-2)/2}(\Lambda r)$$
(5c)

$$= {}_{0}F_{1}\left(;\frac{\nu_{1}}{2};\frac{\Lambda r}{4}\right),\tag{5d}$$

$$E_t^{(\nu_2)}(\theta|\delta) = \sum_{j=0}^{\infty} \frac{\Gamma((j+1+\nu_2)/2)}{\Gamma((1+\nu_2)/2)} \frac{(\sqrt{2\delta}\cos\theta)^j}{j!}$$
(5e)

$$= {}_{1}F_{1}\left(\frac{\nu_{2}+1}{2};\frac{1}{2};\frac{\delta^{2}\cos^{2}\theta}{2}\right) + \sqrt{2}\,\delta\cos\theta\,\frac{\Gamma(\frac{\nu_{2}+2}{2})}{\Gamma(\frac{\nu_{2}+1}{2})}{}_{1}F_{1}\left(\frac{\nu_{2}+2}{2};\frac{3}{2};\frac{\delta^{2}\cos^{2}\theta}{2}\right),$$

$$E_{N}(r|\delta) = \sum_{i=0}^{\infty}\frac{(\delta r)^{j}}{j!} = {}_{0}F_{0}(;;r) = e^{\delta r}$$
(5f)

$$= {}_{0}F_{1}\left(;\frac{1}{2};\frac{\delta^{2}r^{2}}{4}\right) + \delta r {}_{0}F_{1}\left(;\frac{3}{2};\frac{\delta^{2}r^{2}}{4}\right) = \cosh\delta r + \sinh\delta r = e^{\delta r},$$
(5g)

where

$${}_{p}F_{q}(a_{1},\ldots,a_{p};b_{1},\ldots,b_{q};x) = \sum_{n=0}^{\infty} \frac{(a_{1})_{n}\ldots(a_{p})_{n}}{(b_{1})_{n}\ldots(b_{q})_{n}} \frac{x^{n}}{n!}.$$
(6)

stands for the generalized hypergeometric function, and where  $\mathcal{I}^{(\nu_1-2)/2}(\Lambda r)$  stands for the normalized modified Bessel function [16]. The respective first expansions for the multiplicative factors *E* are successively simpler versions of the Maclaurin expansion for the noncentral *F* distribution multiplicative factor  $E_F^{(\nu_1,\nu_2)}$ , such that

$$E_{\chi^2}^{(\nu_1)} \quad \xleftarrow{}_{\nu_2 \to \infty} \quad E_F^{(\nu_1, \nu_2)} \quad \xrightarrow{}_{\nu_1 = 1} \quad E_t^{(\nu_2)} \xrightarrow{}_{\nu_2 \to \infty} \quad E_N, \tag{7}$$

with the understanding that the respective Maclaurin summations involve either only the even integers for the *F* and  $\chi^2$  cases, or both odd and even integers for the *t* and normal cases. It is straightforward to verify that the classical noncentral *t* and *F* distributions converge with the noncentral normal and  $\chi^2$  distributions, respectively, in the limit  $\nu_2 \rightarrow \infty$ . The expressions for the classical noncentral distributions do not readily regroup terms of similar order in their noncentrality parameter, contrarily to the ultraspherical noncentral distributions, which do so, as is discussed next.

## 3. The Ultraspherical Noncentral Distributions

Using geometrical arguments, the ultraspherical noncentral *t*-distribution for the *t*-statistic (2) on the hypersphere  $S^{\nu_2}$  was shown by Le Blanc [10] to be given by

$$v_t^{(\nu_2)}(\theta|\delta) = T_t^{(\nu_2)}(\theta|\delta) \ v_t^{(\nu_2)}(\theta|0), \tag{8}$$

where  $v_t^{(\nu_2)}(\theta|0)$  is identical to the central *t*-distribution  $\rho_t^{(\nu_2)}(\theta|0)$  given in Equation (1), where the multiplicative distribution translation factor  $T_t^{(\nu_2)}(\theta|\delta)$  is given by

$$T_t^{(\nu_2)}(\theta|\delta) = \frac{1 - \cos\theta\cos\theta_\delta}{[1 - 2\cos\theta\cos\theta_\delta + \cos^2\theta_\delta]^{(\nu_2 + 1)/2}},\tag{9}$$

and where

$$\cos\theta_{\delta} = \frac{\delta/\sqrt{\nu_2}}{\sqrt{(\delta/\sqrt{\nu_2})^2 + 1}} = \frac{\delta}{\sqrt{\delta^2 + \nu_2}}, \quad 0 < \theta_{\delta} < \pi, \tag{10}$$

in terms of the noncentrality parameter  $\delta$ ,  $-\infty < \delta < \infty$ . As it turns out, the translation term  $T_t^{(\nu_2)}(\theta|\delta)$  is a generating function for the ultraspherical or, equivalently, Gegenbauer polynomials: redefining the variables so that  $x = \cos \theta$ ,  $z = \cos \theta_{\delta}$ , and  $b = (\nu_2 - 1)/2$ , we have that

$$T_t^{(b)}(x|z) = \frac{1-xz}{[1-2xz+z^2]^{b+1}} = \sum_{n=0}^{\infty} \frac{2b+n}{2b} C_n^{(b)}(x) z^n$$

$$= (1-xz)^{-(2b+1)} {}_2F_1\left(\frac{2b+1}{2}, \frac{2b+2}{2}; b+\frac{1}{2}; -\frac{(1-x^2)z^2}{(1-xz)^2}\right),$$
(11)

where  $C_n^{(b)}(x)$  are Gegenbauer polynomials with the explicit representation [4]

$$C_n^{(b)}(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k (b)_{n-k}}{k! \ (n-2k)!} (2x)^{n-2k},$$
(12)

where the Pochhammer symbol  $(x)_n$  is defined by the equality

$$(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)} = x(x+1)(x+2)\cdots(x+n-1),$$
(13)

where  $\lfloor y \rfloor$ , the floor of y, is given by the lowest integer such that  $y - 1 < \lfloor y \rfloor \leq y$ , and where the last equality given in terms of the hypergeometric function  $_2F_1$  can be deduced from Rainville [17] (Equation (144.8)). The Gegenbauer polynomials are orthogonal:

$$\int_{x=-1}^{1} C_m^{(b)}(x) C_n^{(b)}(x) w_t^{(b)}(x) dx = \delta_{m,n} \|C_n^{(b)}\|^2$$
(14)

with respect to the weight function

$$w_t^{(b)}(x) = \frac{\Gamma(b+1)}{\Gamma(\frac{1}{2})\Gamma(b+\frac{1}{2})} (1-x^2)^{b-1/2},$$
(15)

which is identical to the central *t* distribution (1), except for the change of the variable  $x = \cos \theta$ . With this weight function normalization—which, note, differs from the usual weight function for the Gegenbauer polynomials [4]—the norm of the Gegenbauer polynomials simplifies to

$$\|C_n^{(b)}\|^2 = \frac{b}{b+n} \,\frac{(2b)_n}{n!},\tag{16}$$

with, in particular,  $\|C_0^{(b)}\|^2 = 1$ . Note also that Equation (11) could be used to define a generalization  $T_n^{(b)}(x)$  for the Chebyshev polynomials of the first kind,  $T_n^{(b=0)}(x)$ , with

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

which encompasses the defining equation

$$\frac{1-xz}{[1-2xz+z^2]} = \sum_{n=0}^{\infty} T_n^{(0)}(x) \, z^n = \lim_{b \to 0} \sum_{n=0}^{\infty} \frac{n}{2b} \, C_n^{(\Delta)}(x) \, z^n \tag{18}$$

for the Chebyshev polynomials of the first kind. Setting  $x = r/(2b)^{1/2}$  and  $z = \delta/(2b)^{1/2}$ , the central *t* distribution converges with the normal central distribution

$$v_N(r|0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2}$$
(19)

in the limit  $b \to \infty$ . Applying the same limit to the Gegenbauer polynomials and their generating function (11), we have that

$$\lim_{b \to \infty} T_t^{(b)}(x|z) = \lim_{b \to \infty} \frac{1 - xz}{[1 - 2xz + z^2]^{b+1}}$$
  
= 
$$\lim_{b \to \infty} \sum_{n=0}^{\infty} \frac{2b + n}{2b} C_n^{(b)} \left(\frac{r}{(2b)^{1/2}}\right) \left(\frac{\delta}{(2b)^{1/2}}\right)^n$$
  
= 
$$\sum_{n=0}^{\infty} He_n(r) \frac{\delta^n}{n!} = e^{r\delta - \delta^2/2} = T_N(r|\delta),$$
 (20)

where we have used the limit result [4]

$$\lim_{b \to \infty} \frac{1}{(2b)^{n/2}} C_n^{(b)} \left( \frac{r}{(2b)^{1/2}} \right) = \frac{He_n(r)}{n!}.$$
(21)

The multiplicative factor  $T_N(r|\delta)$  imparts a non-vanishing mean  $\delta$  to the zero mean normal distribution, since

$$T_N(r|\delta) v_N(r|0) = e^{\delta r - \delta^2/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(r-\delta)^2} = v_N(r|\delta),$$
(22)

where one recognizes the generating function  $e^{\delta r - \delta^2/2}$  for the Hermite polynomials  $He_n(r)$ , with the explicit representation [4]

$$He_n(r) = n! \sum_{\ell=0}^{\lfloor n/2 \rfloor} \frac{(-1)^{\ell} r^{n-2\ell}}{2^{\ell} \ell! (n-2\ell)!}.$$
(23)

The latter are orthogonal with respect to their defining weight function

$$w_N(r) = v_N(r|0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2},$$
 (24)

with norm

$$\int_{r=-\infty}^{\infty} He_m(r) He_n(r) w_N(r) dr = \delta_{m,n} n!,$$
(25)

and with  $||He_0||^2 = 1$  in particular.

Using geometrical arguments, the ultraspherical noncentral *F*-distribution for the *F*-statistic (3) was shown by Le Blanc [10] to be given by the integral representation

$$v_F^{(\nu_1, \nu_2)}(\theta|\Lambda) = T_F^{(\nu_1, \nu_2)}(\theta|\Lambda) \, v_F^{(\nu_1, \nu_2)}(\theta|0), \tag{26}$$

where  $v_F^{(\nu_1, \nu_2)}(\theta|0)$  is identical to the central *F* distribution  $\rho_F^{(\nu_1, \nu_2)}(\theta|0)$  given in Equation (1), where the multiplicative distribution translation factor  $T_F^{(\nu_1, \nu_2)}(\theta|\Lambda)$  is given by the integral

$$T_{F}^{(\nu_{1},\nu_{2})}(\theta|\Lambda) = \int_{\psi=0}^{\pi} T_{F}^{(\nu_{1},\nu_{2})}(\theta,\psi|\Lambda) \,\mathrm{d}\psi$$
(27)

with integrand

$$T_F^{(\nu_1, \nu_2)}(\theta, \psi | \Lambda) = \frac{(1 - \cos\theta \cos\psi \cos\theta_\Lambda)}{[1 - 2\cos\theta \cos\psi \cos\theta_\Lambda + \cos^2\theta_\Lambda]^{(\nu_1 + \nu_2)/2}} v_t^{(\nu_1 - 1)}(\psi | 0), \quad (28)$$

and where

$$\cos\theta_{\Lambda} = \sqrt{\Lambda/(\Lambda + \nu_2)}, \qquad 0 \le \Lambda < \infty, \quad 0 \le \cos\theta_{\Lambda} < 1,$$
 (29)

in terms of the noncentral parameter  $\Lambda$ . The special case  $\nu_1 = 1$  is given by

$$v_F^{(\nu_1=1,\nu_2)}(\theta|\Lambda) = \sum_{\cos\theta' = [\cos\theta, -\cos\theta]} \frac{1 - \cos\theta' \cos\theta_\Lambda}{[1 - 2\cos\theta' \cos\theta_\Lambda + \cos^2\theta_\Lambda]^{(\nu_2+1)/2}} v_t^{(\nu_2)}(\theta|0).$$
(30)

Setting  $a = (v_1 - 2)/2$ ,  $b = (v_2 - 2)/2$ ,  $x = \cos \theta$ ,  $\xi = \cos \psi$ ,  $z = \cos \theta_{\Lambda}$ , and making use of the Gegenbauer polynomial-generating function expansion (11) together with the corresponding polynomials' explicit expression (12), the integration in (27) can be carried out: we find

$$T_{F}^{(a,b)}(x|z) = \int_{\xi=-1}^{1} \frac{1-x\xi z}{[1-2x\xi z+z^{2}]^{a+b+2}} w_{t}^{(a)}(\xi) d\xi \qquad (31)$$

$$= \sum_{n=0}^{\infty} \frac{a+b+1+n}{a+b+1} \mathcal{F}_{n}^{(a,b)}(x^{2}) z^{2n},$$

$$= \sum_{n=0}^{\infty} (-1)^{n} \frac{(a+b+2)_{n}}{(a+1)_{n}} P_{n}^{(a,b)}(1-2x^{2}) z^{2n}$$

$$= (1+z^{2})^{-(a+b+2)} {}_{2}F_{1}\left(\frac{1}{2}(a+b+2), \frac{1}{2}(a+b+3); a+1; \frac{4x^{2}z^{2}}{(1+z^{2})^{2}}\right),$$

where the polynomials  $\mathcal{F}_n^{(a,b)}(x^2)$  which one related to the Jacobi polynomials [4], through

$$\mathcal{F}_n^{(a,b)}(x^2) = (-1)^n \frac{(a+b+1)_n}{(a+1)_n} P_n^{(a,b)}(1-2x^2)$$
(32)

are provided with the explicit representation

$$\mathcal{F}_{n}^{(a,b)}(x^{2}) = \sum_{k=0}^{n} \frac{(-1)^{k}}{k!} \frac{(a+b+1)_{2n-k}}{(a+1)_{n-k}} \frac{(x^{2})^{n-k}}{(n-k)!},$$
(33)

and where the last equality—deduced from Rainville [17] (Equation (132.10))—provides us with a generating function for the polynomials in terms of the hypergeometric function  $_2F_1$ . The  $\mathcal{F}_n^{(a,b)}(x^2)$  polynomials are orthogonal with respect to the weight function

$$w_F^{(a,b)}(x) = 2 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} (x^2)^{a+\frac{1}{2}} (1-x^2)^b, \quad 0 \le x \le 1,$$
(34)

which is identical to the central *F* distribution (1), except for the change of the variable  $x = \cos \theta$ , with norm

$$\|\mathcal{F}_{n}^{(a,b)}\|^{2} = \frac{a+b+1}{a+b+1+2n} \frac{(a+b+1)_{n}(b+1)_{n}}{(a+1)_{n} n!},$$
(35)

and with  $\|\mathcal{F}_0^{(a,b)}\|^2 = 1$  in particular. Equation (31) can be verified to be valid for the special case (30) with a = -1/2 ( $\nu_1 = 1$ ). In the limit  $\nu_2 \to \infty$ , the central *F* distribution converges with the central  $\chi^2$  distribution

$$v_{\chi^2}^{(\nu_1)}(r|0) = \frac{1}{2^{\nu_1/2}\Gamma(\frac{\nu_1}{2})} r^{(\nu_1-2)/2} e^{-r/2}, \quad 0 \le r < \infty,$$
(36)

which, with  $a = (v_1 - 2)/2$ , can be used as a normalized defining weight function,

$$w_{\chi^2}^{(a)}(r) = \frac{1}{2^{a+1}} \frac{1}{\Gamma(a+1)} r^a e^{-r/2}, \quad a > -1,$$
(37)

for the Laguerre orthogonal polynomial family. The Laguerre polynomials can be given the explicit representation [4]

$$L_n^{(a)}(y) = \sum_{\ell=0}^n (-1)^\ell \; \frac{(a+\ell+1)_{n-\ell}}{(n-\ell)!} \; \frac{y^\ell}{\ell!}, \quad 0 \le y < \infty.$$
(38)

Since Equation (32) provides the noncentral *F* distribution translation factor  $T_F^{(a,b)}(x|z)$  with an expansion in terms of the Jacobi polynomials  $P_n^{(a,b)}(1-2x^2)$ , and since the limit result

$$\lim_{b \to \infty} P_n^{(a,b)} \left( 1 - \frac{2y}{b} \right) = L_n^{(a)}(y) \tag{39}$$

holds [4], one verifies with  $x^2 = r/2b$  and  $z^2 = \Lambda/2b$  that

$$\lim_{b \to \infty} T_F^{(a,b)}(x|z) = T_{\chi^2}^{(a)}(r|\Lambda) = \mathcal{I}^{(a)}(\Lambda r) e^{-\Lambda/2} = {}_0F_1\left(;a+1;\frac{\Lambda r}{4}\right) e^{-\Lambda/2} \\ = \sum_{n=0}^{\infty} \frac{(-1)^n}{(a+1)_n} L_n^{(a)}(r/2) (\Lambda/2)^n,$$
(40)

which is a lesser-known expression for the noncentral  $\chi^2$  distribution translation factor first derived by Tiku [18]. Under normalization (37) for their defining weight factor, the norm of the Laguerre polynomials  $L_n^{(a)}(r/2)$  is given by

$$\int_{r=0}^{\infty} L_m^{(a)}(r/2) L_n^{(a)}(r/2) w_{\chi^2}^{(a)}(r) dr = \delta_{m,n} \|L_n^{(a)}\|^2 \delta_{m,n} = \frac{(a+1)_n}{n!}$$
(41)

with, in particular,  $||L_0^{(a)}||^2 = 1$ .

To summarize, the four ultraspherical noncentral F,  $\chi^2$ , t, and the normal N distributions are given by the product of translation factors T, which take the form of generating functions for specific orthogonal polynomial families, times the corresponding normalized central distributions which also stand for the corresponding polynomial family-defining weight functions. Thus, we have that

$$v_F^{(a,b)}(x|z) = T_F^{(a,b)}(x|z) w_F^{(a,b)}(x), \quad 0 \le x < 1, \quad 0 \le z < 1,$$
 (42a)

$$v_{\chi^2}^{(a)}(r|\Lambda) = T_{\chi^2}^{(a)}(r|\Lambda) \ w_{\chi^2}^{(a)}(r), \quad 0 \le r < \infty, \quad 0 \le \Lambda < \infty,$$
(42b)

$$v_t^{(b)}(x|z) = T_t^{(b)}(x|z) \ w_t^{(b)}(x), \quad -1 < x < 1, \quad -1 < z < 1,$$
(42c)

$$w_N(r|\delta) = T_N(r|\delta) w_N(r), \quad -\infty < r < \infty, \quad -\infty < \delta < \infty, \tag{42d}$$

with the normalized central distributions

$$w_F^{(a,b)}(x) = 2 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} (x^2)^{a+\frac{1}{2}} (1-x^2)^b,$$
(43a)

$$w_{\chi^2}^{(a)}(r) = \frac{1}{2^{a+1}} \frac{1}{\Gamma(a+1)} r^a e^{-r/2},$$
 (43b)

$$w_t^{(b)}(x) = \frac{\Gamma(b+1)}{\Gamma(\frac{1}{2})\Gamma(b+\frac{1}{2})} (1-x^2)^{b-1/2}$$
(43c)

$$w_N(r) = v_N(r|0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}r^2},$$
 (43d)

providing the polynomial family-defining weights, and with the corresponding generating functions and orthogonal polynomial expansions given by

$$T_{F}^{(a,b)}(x|z) = (1+z^{2})^{-(a+b+2)}{}_{2}F_{1}\left(\frac{1}{2}(a+b+2), \frac{1}{2}(a+b+3); a+1; \frac{4x^{2}z^{2}}{(1+z^{2})^{2}}\right)$$
(44a)  
$$-\sum_{k=0}^{\infty} (-1)^{n} \frac{(a+b+2)_{n}}{2} p_{1}^{(a,b)}(1-2x^{2}) z^{2n} - \sum_{k=0}^{\infty} \frac{a+b+1+n}{2} r_{1}^{(a,b)}(x^{2}) z^{2n}$$

$$=\sum_{n=0}^{\infty} (-1)^n \frac{(1-2)^n}{(a+1)_n} P_n^{(a,b)} (1-2x^2) z^{2n} = \sum_{n=0}^{\infty} \frac{(1-2)^n}{(a+b+1)^n} \mathcal{F}_n^{(a,b)} (x^2) z^{2n},$$

$$T_{\chi^2}^{(a)}(r|\Lambda) = {}_0F_1\left(;a+1;\frac{\Lambda r}{4}\right) e^{-\Lambda/2} = \mathcal{I}^{(a)}(\Lambda r) e^{-\Lambda/2} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(a+1)_n} L_n^{(a)}(r/2) (\Lambda/2)^n,$$
(44b)

$$T_t^{(b)}(x|z) = (1 - xz)^{-(2b+1)} {}_2F_1\left(\frac{2b+1}{2}, \frac{2b+2}{2}; b + \frac{1}{2}; -\frac{(1 - x^2)z^2}{(1 - xz)^2}\right)$$
(44c)

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

$$(r|\delta) = {}_{2}F_{2}(\cdots \delta r)e^{-\delta^2/2} - e^{r\delta - \delta^2/2} - \sum_{n=0}^{\infty} He_n(r) \frac{\delta^n}{2r}$$
(44d)

$$T_N(r|\delta) = {}_0F_0(;;\delta r)e^{-\delta^2/2} = e^{r\delta - \delta^2/2} = \sum_{n=0}^{\infty} He_n(r) \frac{\delta^n}{n!}.$$
(44d)

We conclude this section by stressing the fact that the random variable space for the ultraspherical noncentral distributions are translated hyperspheres rather than translated normal distributions, as is assumed for the classical noncentral distributions. The tide model developed in Section 6 provides a concrete example of such a distribution on a translated sphere. As more extensively argued in [15], the ultraspherical and classical noncentral *F* and *t* distributions correspond to projections of translated hyperspheres and translated normal distributions, respectively; are identical in their central distributions when their noncentrality parameters are zero; and converge in high-dimensional spaces, but diverge in low-dimension spaces and for large noncentrality parameters. See Figure 1. 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 [3], allowing one to use the ultraspherical distributions as surrogates for the classical noncentral *t* and *F* distributions in high-dimensional spaces.



**Figure 1.** Symmetrized Kullback–Leibler divergence (**a**) between the hyperspherical  $v_t^{(\nu_2)}(\theta|\delta)$  and the classical  $\rho_t^{(\nu_2)}(\theta|\delta)$  noncentral *t*—distributions in the upper-left-corner plot, and (**b**–**f**) between the hyperspherical  $v_F^{(\nu_1,\nu_2)}(\theta|\Lambda)$  and the classical  $\rho_F^{(\nu_1,\nu_2)}(\theta|\Lambda)$  noncentral *F* distributions for  $v_1 = (2, ..., 6)$ , respectively, for all the other plots. The ultraspherical and classical noncentral *t* and *F* distributions correspond to projections of translated uniform distributions on unit radius hyperspheres and translated normal distributions, respectively; are identical in their central distributions when their noncentrality parameters  $\delta$  or  $\Lambda$  vanish; converge in high-dimensional (large degree-of-freedom  $v_2$ ) spaces, but diverge in low-dimension spaces and for large noncentrality  $\delta$  or  $\Lambda$  parameters. The ultraspherical noncentral distributions can be used as surrogates for the classical noncentral *t* and *F* distributions in high-dimensional spaces. See the text for details.

## 4. Entropic Convex Duals Expansion in Terms of Orthogonal Polynomials

Bayes–Jaynes–Gibbs data-constrained maximal entropy priors—simply designated as Gibbs priors in the following—can be objectively computed for dense datasets. See Le Blanc [8] for an extensive review on the subject, from which we recall that the Bayes–Laplace prior and posterior update rules are rooted in the convex geometry of Shannon's entropy function, with the Kullback–Leibler relative entropy being a Bregman divergence defined in terms of the former. The Gibbs priors can be formally expressed as

$$\pi(r_o) = \frac{1}{Z(\lambda)} \exp\left(\int_{\mathcal{R}} \lambda(r) v(r|r_o) \, \mathrm{d}r\right),\tag{45}$$

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

$$Z(\lambda) = \int_{\mathcal{R}_o} \exp\left(\int_{\mathcal{R}} \lambda(r) v(r|r_o) \, \mathrm{d}r\right) \mathrm{d}r_o, \tag{46}$$

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

$$\inf_{\lambda} G_{\rho}(\lambda) = \inf_{\lambda} \left[ \log Z(\lambda) - \int_{\mathcal{R}} \lambda(r) v(r) \, \mathrm{d}r \right], \tag{47}$$

which, by convex duality, corresponds to Jaynes data-constrained maximal entropy. The corresponding Gibbs–Jaynes model for a generic empirical density v(r) is given by

$$v(r) = \int_{\mathcal{R}_o} v(r|r_o) \ \pi(r_o) \ dr_o = v(r|0) \ \int_{\mathcal{R}_o} T(r|r_o) \ \pi(r_o) \ dr_o = v(r|0) \ BF(r).$$
(48)

As posed, solving for the entropic convex dual function  $\lambda(r)$  in (47) requires one to compute its value across the entire support domain of the empirical distribution v(r), a task which can be expensive in computing terms. Now, since the ultraspherical noncentral distribution translation factors  $T(r|r_0)$  are generating functions for orthogonal polynomial families, and since the corresponding central distributions v(r|0) are the family-defining weight functions  $w(r) \equiv v(r|0)$ , one can rewrite the unconstrained minimization problem in simpler terms. Indeed, the exponentiated term in the partition function (46) can be rewritten as

$$\int_{\mathcal{R}} \lambda(r) v(r|r_{o}) dr = \int_{\mathcal{R}} \lambda(r) T(r|r_{o}) w(r) dr = \int_{\mathcal{R}} \lambda(r) \left[ \sum_{n=0}^{\infty} c_{n} P_{n}(r) r_{o}^{n} \right] w(r) dr$$

$$= \sum_{n=0}^{\infty} c_{n} \left[ \int_{\mathcal{R}} \lambda(r) P_{n}(r) w(r) dr \right] r_{o}^{n} = \sum_{n=0}^{\infty} c_{n} \lambda_{n} ||P_{n}||^{2} r_{o}^{n}$$

$$= \sum_{n=0}^{\infty} \tilde{\lambda}_{n} r_{o}^{n} \text{ with } \tilde{\lambda}_{n} = c_{n} \lambda_{n} ||P_{n}||^{2}, \qquad (49)$$

where the polynomials  $P_n$ , with their respective multiplicative coefficients  $c_n$ , are listed in (44), and where the  $\lambda_n$  are the entropic convex dual expansion coefficients on family-wise orthogonal polynomials that remain to be determined. Similarly, the additive constraint term in (47) can be rewritten as

$$\int_{\mathcal{R}} \lambda(r) v(r) dr = \sum_{n=0}^{\infty} \lambda_n \int_{\mathcal{R}} P_n(r) v(r) dr = \sum_{n=0}^{\infty} \lambda_n \int_{\mathcal{R}} P_n(r) BF(r) w(r) dr = \sum_{n=0}^{\infty} \tilde{\lambda}_n \tilde{\beta}_n,$$
(50)

where v(r) is the empirical density to be modeled, and where

$$\tilde{\beta}_n = \frac{1}{c_n \|P_n\|^2} \int_{\mathcal{R}} P_n(r) BF(r) w(r) dr,$$
(51)

with, in particular,  $\tilde{\beta}_0 = 1$ . Within the division by the factor  $c_n$ , the latter equality is the Bayes factor BF(r) expansion on the orthogonal polynomials in random sample space. The unconstrained optimization problem can thus be reformulated as

$$\inf_{\{\tilde{\lambda}_n\}} \left[ \log \left( \int_{\mathcal{R}_o} \exp(\sum_{n=0}^{\infty} \tilde{\lambda}_n r_o^n) \, \mathrm{d}r_o \right) - \sum_{n=0}^{\infty} \tilde{\lambda}_n \tilde{\beta}_n \right]$$
(52)

in terms of orthogonal polynomial expansion coefficient sets  $\{\lambda_n\}_{n=0}^{\infty}$  for the continuous entropic convex dual function  $\lambda(r)$ , sets which can be restricted to a small finite number of coefficients, as can be assessed by the Kullback–Leibler divergence between the empirical

density and its model in terms of orthogonal polynomials. At the minimum of the Gibbs potential (52), one has the simple condition

$$\frac{\int_{\mathcal{R}_o} \exp(\sum_n \tilde{\lambda}_n r_o^n) r_o^n \, \mathrm{d}r_o}{\int_{\mathcal{R}_o} \exp(\sum_n \tilde{\lambda}_n r_o^n) \, \mathrm{d}r_o} = \int_{\mathcal{R}_o} \pi(r_o) r_o^n \, \mathrm{d}r_o = \tilde{\beta}_n, \tag{53}$$

which states that the  $n^{\text{th}}$  moment of the noncentrality parameter  $r_o$ , as weighted by the Gibbs prior

$$\pi(r_o) = \frac{\exp(\sum_n \lambda_n r_o^n)}{\int_{\mathcal{R}_o} \exp(\sum_n \tilde{\lambda}_n r_o^n) \, \mathrm{d}r_o'}$$
(54)

is, within the factor  $c_n$ , equal to the  $n^{\text{th}}$  coefficient of the Bayes factor expansion (51) on the orthogonal polynomial basis in random sample space. In practical terms, the determination of the Gibbs prior in terms of a small number of polynomial expansion coefficients  $\tilde{\lambda}_n$  for the entropic dual convex results in a substantial reduction in the computing time needed to find the Gibbs potential minimum.

Once obtained, the Gibbs prior can be used to modelize the empirical density v(r) according to Equation (48), in which the central distribution v(r|0) stands for the weight function w(r) for the ultraspherical distributions listed in (42). A generic density v(r) can thus be expanded as

$$v(r) = \int_{\mathcal{R}_{o}} v(r|r_{o}) \pi(r_{o}) dr_{o} = w(r) \int_{\mathcal{R}_{o}} T(r|r_{o}) \pi(r_{o}) dr_{o}$$
(55)  
$$= w(r) \int_{\mathcal{R}_{o}} \left[ \sum_{n=0}^{\infty} c_{n} P_{n}(r) r_{o}^{n} \right] \pi(r_{o}) dr_{o} = w(r) \sum_{n=0}^{\infty} c_{n} \left[ \int_{\mathcal{R}_{o}} r_{o}^{n} \pi(r_{o}) dr_{o} \right] P_{n}(r)$$
$$= w(r) \sum_{n=0}^{\infty} c_{n} \tilde{\beta}_{n} P_{n}(r) = w(r) \sum_{n=0}^{\infty} \left[ \frac{1}{\|P_{n}\|^{2}} \int_{\mathcal{R}} P_{n}(r') BF(r') w(r') dr' \right] P_{n}(r)$$

where we have invoked the optimality condition (53) in going from the second to the third line. This sequence of equalities provides us with two alternative ways,

$$BF(r) = \frac{v(r)}{w(r)} = \int_{\mathcal{R}_o} T(r|r_o) \, \pi(r_o) \, dr_o \tag{56a}$$

$$=\sum_{n=0}^{\infty} c_n \left[ \int_{\mathcal{R}_o} r_o^n \, \pi(r_o) \, \mathrm{d}r_o \right] \, P_n(r) = \sum_{n=0}^{\infty} \left[ \frac{1}{\|P_n\|^2} \int_{\mathcal{R}} P_n(r') \, BF(r') \, w(r') \, \mathrm{d}r' \right] \, P_n(r), \tag{56b}$$

to model the Bayes factors and the empirical densities: either one determines the Gibbs prior  $\pi(r_o)$  via the optimization problem (52) and uses it to weigh the analytical orthogonal polynomial-generating functions  $T(r|r_o)$  listed in (44); or, one directly determines the prior moments in parametric space in terms of the Bayes factor BF(r) expansion coefficients in random variable space, as per the last equality above. The latter strategy amounts to the construction of a probability distribution in terms of its moments [12]. In that respect, recall the Hausdorff moment problem, which stipulates that the collection of all moments of a probability distribution on a bounded interval uniquely determine the distribution [19], and the Hamburger moment problem, which considers the uniqueness of solutions for the same problem on the unbounded real line [20].

Numerical exploration indicates that the convergence of the polynomial expansion for the ultraspherical noncentral *t* and *F* distributions' generating functions is affected by the Gibbs phenomenon at the extremities of their finite noncentrality parameter ranges [21]. For this reason, when considering the noncentral *t* and *F* distributions, one can either solve for the Gibbs prior to weigh the corresponding analytic generating functions (44) in constructing the models, or use the large *b* limit for the distributions which converge with the noncentral normal *N* and  $\chi^2$  distributions, respectively, together with the associated Hermite and Laguerre orthogonal polynomial families, to directly compute the prior

moments in parametric space in terms of the Bayes factor expansion coefficients in random variable space, as according to (56). Both strategies are found to be unaffected by the Gibbs effect, as exemplified in the next section.

## 5. Genomics Examples

We first revisit, in this section, the NCBI Gene Expression Omnibus head and neck squamous cell carcinoma microarray dataset produced by [22], pertaining to 22 paired samples (N = 44,  $\nu_2 = 42$ , b = 41/2) of normal versus cancerous tissues, and interrogating 11,302 genes via 2-sample t-tests. In Le Blanc [10], a maximal entropy prior equal to the central ultraspherical t distribution (1) was postulated. In Le Blanc [8], the entropic dual convex  $\lambda(x)$  was computed on the full random variable range -1 < x < 1. Here, the Bayes factor is provided with two models. The first model  $v(x) = w_t(x) \int T(x|z)\pi(z)dz$  is given in terms of the Gibbs prior  $\pi(z)$  (54) and the analytical Gegenbauer polynomial-generating function T(x|z) (11). We assessed the computational gains in our determination of the entropic convex dual  $\lambda(x)$  via Equation (52): with the empirical distribution v(x) binned in 200 bins, and using the MATLAB<sup>®</sup> optimization subroutine fminunc in a 64-bit Windows environment on a PC with an Intel(R) Core(TM) i9-9900K CPU @ 3.60 GHz processor and 64 GB of RAM, the determination times for the entropic convex dual  $\lambda(x)$  in terms of a small number of Gegenbauer polynomial coefficients were curtailed by a factor ranging from 100 to 1000, compared to previous determination times of  $\lambda(x)$  on the full random variable range -1 < x < 1, with elapsed times as short as a few tenths of a second. In the second model, according to (56) and with the Hermite polynomials standing for the Gegenbauer polynomials in the large b limit, the Bayes factor BF(r) expansion in random variable space on the Hermite polynomials provides us with the Gibbs prior moments  $\int \pi(\delta) \delta^n d\delta$  in parametric space. As can be seen in Figure 2, the convergence of the two models with the empirical distribution is rapidly achieved with ten coefficients or less.

Next, we study a genome-wide association study (GWAS) dataset. 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. Such studies 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 2244 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 the UK Biobank [23]. A logistic regression model was used for each of the 4,380,209 SNPs individually tested for statistical significance. We computed the empirical density of all the statistical test-associated  $\chi^2_{\nu_1=1}$ *r*-statistics, and used it to compute the first eight terms of the direct polynomial expansion (56) on Laguerre polynomials for the Bayes Factor BF(p(r)). As illustrated in Figure 3, the accrual of the successive polynomial expansion terms allows for an incrementally better fit of the *p*-value empirical density, which strongly deviates from the NHST null hypothesis U(0,1) in the low p-value range where associations are detected. In the Bayesian framework, NHST statistical significance is replaced by the strength of Bayesian evidence, as assessed by magnitude of BF(p), which, in turn, allows for the computation of a local false discovery rate [11]

$$fdr(p) = 1/(1 + BF(p))$$
 (57)

as is also illustrated in Figure 3.



**Figure 2.** Empirical random space densities and NHST *p*-value densities modelization for the head and neck cancer dataset. The upper panels illustrate the Jaynes–Gibbs model  $v(x) = w_t(x) \int T(x|z)\pi(z)dz$ , as provided by the Gibbs prior (54) and the analytical generating function (11) for the Gegenbauer polynomials. (a) Upper left-hand panel: convergence of the Jaynes–Gibbs model with the empirical density  $v(x) = w_t(x) BF(x)$ . (b) Upper right-hand panel: corresponding NHST *t*-test *p*-value model densities, as modeled by the Bayes factor BF(p). As can be observed, convergence on the empirical densities is rapidly achieved with the expansion of the entropic convex dual  $\lambda(x)$  on a small number *n* of Gegenbauer polynomials. In the lower panels, with the Hermite polynomials standing for the Gegenbauer polynomials in the large *b* limit (21), the Bayes factor BF(r) expansion coefficients in random variable space directly provides the Gibbs prior moments  $\int \pi(\delta)\delta^n d\delta$  in parametric space, as according to (56). (c) Lower left-hand panel: cumulative orthogonal polynomial expansion of the Bayes factor BF(r) in random variable space. (d) Lower right-hand panel: corresponding NHST central normal distribution *p*-value densities as modeled by the Bayes factor BF(p). As can be observed, convergence on the empirical normal distribution *p*-value densities is rapidly achieved with a small number *n* of low-order Hermite polynomials.



**Figure 3.** Bayes Factor BF(p) modeling a NHST *p*-value distribution from a genome-wide association study (GWAS) dataset. The GWAS compared 2244 critically ill patients with COVID-19 with 3 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. (a) Left panel: Accrual of the successive Laguerre polynomial expansion terms in Equation (56) for the Bayes factor demonstrating an incrementally better fit of the *p*-value empirical density which strongly deviates from the NHST null hypothesis U(0,1) in the low *p*-value range. (b) Right panel: Local false discovery rate fdr(p) = 1/(1 + BF(p)). The Bayesian-based fdr crosses the 0.01 threshold (i.e., a fdr of 1%) when the NHST *p*-value reaches about  $10^{-7}$  ( $-\log_{10}(p) = 7$ ), in close concordance with the threshold of significance of  $5 \times 10^{-8}$  ( $-\log_{10}(p) = 7.3$ ) chosen by the authors.

# 6. Geophysics Examples

We begin this section by modeling, in Figure 4, Earth's above-sea-level emerging land/ice latitudinal density [24]. Because an ideal Earth surface can be described by the sphere  $S^{v_2=2}$ , we have that the parameter  $b = (v_2 - 1)/2 = 1/2$ , with the weight factor  $w_t^{(1/2)}(x)$  in (15) reducing to the constant 1/2. As a consequence, Equation (55) is equivalent to a standard expansion on Legendre polynomials—equivalently, Gegenbauer  $C_n^{(1/2)}$  polynomials—except for the fact that our convention regarding the central distribution normalization  $w_t^{(1/2)}(x) = 1/2$  (rather than the usual weight w(x) = 1) on the span [-1,1] ensures that expansion (55) defines a normalized density.



**Figure 4.** Earth's emerging land/ice latitudinal density. Orthogonal Legendre  $P_n$  polynomial (Gegenbaeur  $C_n^{(1/2)}$  polynomial) modeling of Earth's emerging land/ice masses' latitudinal density. (a) Left panel: orthogonal Legendre  $P_n$  polynomial modeling (55) of Earth's emerging land/ice latitudinal density as expanded on the first 30 Legendre polynomials. (b) Right panel: Kullback–Leibler divergence between the empirical density and the model as a function of the number of orthogonal Legendre polynomials accrued.

We conclude with an example in which the translating factor (9) for the ultraspherical noncentral *t* distribution is kept as-is, instead of using its expansion (11) in terms of Gegenbauer polynomials, to argue that the noncentral spherical *t* distribution (8) on  $S^2$  can be used to describe the geometry of the gravity forces of a simple tide model. Consider Figure 5, which describes the gravitational pull of an ideal Moon of mass *m* on the thin water layer covering an ideal Earth of mass *M* and radius *R* at distance *D* from the Moon. The gravitational tidal force T per unit mass on point *s* on the Earth's surface is given by

$$\mathcal{T} = \frac{Gm}{(D - R\cos\theta)^2 + R^2\sin^2\theta} = \frac{Gm}{D^2 - 2DR\cos\theta + R^2}.$$
(58)

The horizontal and vertical components of that force are given by

$$\mathcal{T}_h = +\frac{Gm}{D^2 - 2DR\cos\theta + R^2} \times \cos\varphi = +Gm\frac{D - R\cos\theta}{[D^2 - 2DR\cos\theta + R^2]^{3/2}},$$
 (59)

$$\mathcal{T}_{v} = -\frac{Gm}{D^{2} - 2DR\cos\theta + R^{2}} \times \sin\varphi = -Gm\frac{R\sin\theta}{[D^{2} - 2DR\cos\theta + R^{2}]^{3/2}}, \quad (60)$$

respectively. Setting our physical units so that  $Gm/D^2 = 1$ , and defining  $\cos \theta_{\delta} = R/D$ , we have that

$$\mathcal{T}_{h} = \frac{1 - \cos\theta\cos\theta_{\delta}}{[1 - 2\cos\theta\cos\theta_{\delta} + \cos^{2}\theta_{\delta}]^{3/2}} = T_{t}^{(\nu_{2}=2)}(\theta|\delta), \tag{61}$$

$$\mathcal{T}_{v} = \frac{-\sin\theta}{[1 - 2\cos\theta\cos\theta_{\delta} + \cos^{2}\theta_{\delta}]^{3/2}}.$$
(62)



**Figure 5.** Tide geometry. Idealized model describing the gravitational tidal pull of an ideal Moon of mass *m* on a thin water layer covering an ideal Earth of mass *M* and radius *R* at distance *D* from the Moon. The horizontal tidal force is given by the modular translation factor  $T_t^{(\nu 2=2)}(\theta|\delta)$  (9), defining on  $S^2$  the noncentral spherical distribution  $v_t^{(\nu_2=2)}(\theta|\delta)$  (8) with  $\cos \theta_{\delta} = R/D$ , minus a factor of one, accounting for the centrifugal force induced by the Moon–Earth system revolving around its center of gravity. The equatorial bulges are not symmetric in this purely geometrical model. The noncentrality parameter  $\cos \theta_{\delta}$  is set to an unrealistic value of 0.2 to enhance the visualization of the geometrical distribution of the tidal forces.

The horizontal tidal force component  $\mathcal{T}_h$  is thus simply given by the noncentral spherical *t* distribution translation factor  $T_t^{(\nu_2=2)}(\theta|\delta)$ , as provided by Equation (4). One can integrate this horizontal tidal force over the entire spherical shell by weighing it with the central spherical distribution  $v_t^{(\nu_2=2)}(\theta|0)$  to account for the spherical geometry of the Earth. Since the integrand to this integral corresponds to the noncentral spherical *t* distribution  $v_t^{(\nu_2=2)}(\theta|\delta) = T_t^{(\nu_2=2)}(\theta|\delta) v_t^{(\nu_2=2)}(\theta|0)$  on  $S^2$ , the tidal force integrates to one in our unit system. In order for the system to be stationary, an opposing force must be opposed to this integrated tidal force of one. The inertial centrifugal force originating from the joint rotation of the Moon and Earth around their center of mass plays such a role. It is evaluated to be given by  $-Gm/D^2 = -1$  in our unit system [25], which we add to the normalized horizontal tidal force  $\mathcal{T}_h$  above. We have plotted the resulting force field in Figure 5, with the noncentrality parameter  $\cos \theta_{\delta}$  set to an unrealistic value of 0.2 to enhance the visualization of the geometrical distribution of the tidal forces. Unlike most illustrative tide diagrams with symmetrical bulges found in the literature, our tide diagram, together with its unsymmetrical equatorial water bulges, provides a more accurate visual depiction of the expected asymmetry of the tidal forces resulting from purely geometrical considerations.

#### 7. Discussion

We have seen that the univariate noncentral distributions can be constructed in a modular fashion by multiplying their central distributions with specific translation factors. Using geometrical arguments, we have found that the translation factors for the ultraspherical noncentral *t*, normal, *F*, and  $\chi^2$  distributions stand for generating functions for the Gegenbauer, Hermite, Jacobi, and Laguerre polynomial families, respectively, with their central distributions standing for the corresponding polynomial family-defining weights. These developments clearly link four of the most important classical continuous probability distributions with the powerful orthogonal polynomial formalism. To the best of our knowledge, the derivation of these translation factors and their identification as orthogonal polynomial family-generating functions has not been carried out before.

Jaynes' maximal entropy prior is obtained through the unconstrained minimization of the Gibbs potential. In parametric Bayesian inference, the formal expression for the Gibbs potential comprises an integral of the product of the empirical distribution's entropic convex dual  $\lambda(r)$  times the parametric kernel, or, equivalently, the likelihood function  $\rho(r|r_o)$ . In the case of the ultraspherical noncentral distributions, the latter integral yields discretized expansion coefficients of the entropic convex duals on orthogonal polynomial bases. The determination of the entropic convex duals is thus reduced to the much simpler and computationally economical determination of a few low-order orthogonal polynomial coefficients. By invoking the moment problem and the duality principle, prior moments in parametric space are equated with Bayes factors expansion coefficients over orthogonal polynomial bases in random variable space. To the best of our knowledge, the expansion and discretization of convex duals over orthogonal polynomial bases has not been proposed before. In an approach which bears some similarities to our work, Alibrandi and Ricciardi [26] proposed a moment-based approach, the use of a discretized kernel set  $\{\rho(r|r_i)\}_{i=1}^N$ , and the use of Jaynes' maximal entropy principle to determine the set's weighting factors  $\{\pi_i \mid \sum_{i=1}^N \pi_i = 1\}$ . While their kernel discretization procedure is an ad hoc procedure, our kernel discretization is principled since it is based on the powerful orthogonal polynomial formalism.

The machine learning community has begun to exploit the classical orthogonal polynomial formalism. A non-exhaustive review identifies the use of orthogonal polynomials in optical character recognition [27], in support vector machine kernel construction [28], and in polynomial-based iteration methods for symmetric linear systems [29,30]. In a similar vein, we hope to contribute the present formalism to both the statistical and the machine learning communities.

Funding: This research received no external funding.

**Data Availability Statement:** For the head and neck cancer microarray dataset, please refer to [31]. The COVID-19 GWAS dataset was accessed at https://genomicc.org/data, accessed on 3 January 2022. For Earth's emerging land/ice latitudinal density, please refer to [32].

Conflicts of Interest: The author declares no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	analysis of variance
dof	degrees of freedom
FDR	false discovery rate
GWAS	genome-wide association study
NHST	null-hypothesis statistical testing
SNP	single-nucleotide polymorphism

#### References

- 1. Wikipedia Contributors. Effect Size—Wikipedia, The Free Encyclopedia. 2022. Available online: https://en.wikipedia.org/wiki/ Effect\_size (accessed on 7 March 2022).
- Saville, D.J.; Wood, G.R. Statistical Methods: A Geometric Primer; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1996.
- 3. Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science;* Cambridge University Press: Cambridge, UK, 2018; Volume 47.
- 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. NIST Digital Library of Mathematical Functions. 2021. Available online: http://dlmf.nist.gov/ (accessed on 15 September 2021).
- 5. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620. [CrossRef]
- 6. Jaynes, E.T. Information theory and statistical mechanics. II. Phys. Rev. 1957, 108, 171. [CrossRef]
- 7. Pressé, S.; Ghosh, K.; Lee, J.; Dill, K.A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115. [CrossRef]
- 8. Le Blanc, R. Entropic convex duality in the determination of data-constrained kernel-based Bayes-Jaynes priors. *J. Convex Anal.* **2022**, *29*, 623–647.
- 9. Ismail, M.; Ismail, M.E.; van Assche, W. *Classical and Quantum Orthogonal Polynomials in One Variable*; Cambridge University Press: Cambridge, UK, 2005; Volume 13.
- 10. Le Blanc, R. Bayesian Analysis on a Noncentral Fisher–Student's Hypersphere. Am. Stat. 2019, 73, 126–140. [CrossRef]
- 11. Stephens, M. False discovery rates: A new deal. *Biostatistics* 2017, 18, 275–294. [CrossRef] [PubMed]
- 12. Schmüdgen, K. The Moment Problem; Springer: Berlin/Heidelberg, Germany, 2017; Volume 9.
- 13. Rockafellar, R.T. Convex Analysis; Princeton University Press: Princeton, NJ, USA, 2015.
- 14. Walck, C. Hand-book on statistical distributions for experimentalists. Univ. Stockh. 2007, 10, 96-01.
- 15. Le Blanc, R. Noncentral univariate distributions. 2022, to be submitted.
- 16. András, S.; Baricz, Á. Properties of the probability density function of the non-central chi-squared distribution. *J. Math. Anal. Appl.* **2008**, *346*, 395–402. [CrossRef]
- 17. Rainville, E.D. Special Functions; The Macmillan Company: New York, NY, USA, 1960; Volume 5.
- 18. Tiku, M. Laguerre series forms of non-central  $\chi^2$  and *F* distributions. *Biometrika* **1965**, *52*, 415–427. [PubMed]
- 19. Wikipedia Contributors. Hausdorff Moment Problem—Wikipedia, The Free Encyclopedia. 2021. Available online: https://en.wikipedia.org/w/index.php?title=Hausdorff\_moment\_problem&oldid=1016911791 (accessed on 27 February 2022).
- 20. Wikipedia Contributors. Hamburger Moment Problem—Wikipedia, The Free Encyclopedia. 2021. Available online: https://en.wikipedia.org/w/index.php?title=Hamburger\_moment\_problem&oldid=1060022764 (accessed on 27 February 2022).
- 21. Shizgal, B.D.; Jung, J.H. Towards the resolution of the Gibbs phenomena. J. Comput. Appl. Math. 2003, 161, 41–65. [CrossRef]
- 22. Kuriakose, M.; Chen, W.; He, Z.; Sikora, A.; Zhang, P.; Zhang, Z.; Qiu, W.; Hsu, D.; McMunn-Coffran, C.; Brown, S.; et al. Selection and validation of differentially expressed genes in head and neck cancer. *Cell. Mol. Life Sci.* **2004**, *61*, 1372–1383. [CrossRef]
- 23. Pairo-Castineira, E.; Clohisey, S.; Klaric, L.; Bretherick, A.D.; Rawlik, K.; Pasko, D.; Walker, S.; Parkinson, N.; Fourman, M.H.; Russell, C.D.; et al. Genetic mechanisms of critical illness in COVID-19. *Nature* **2021**, *591*, 92–98. [CrossRef]
- 24. Amante, C.; Eakins, B.W. *ETOPO1 Global Relief Model Converted to PanMap Layer Format*; NOAA-National Geophysical Data Center: Boulder CO, USA, 2009. [CrossRef]
- 25. Matsuda, T.; Isaka, H.; Boffin, H.M. Confusion around the tidal force and the centrifugal force. arXiv 2015, arXiv:1506.04085.
- Alibrandi, U.; Ricciardi, G. Efficient evaluation of the pdf of a random variable through the kernel density maximum entropy approach. *Int. J. Numer. Methods Eng.* 2008, 75, 1511–1548. [CrossRef]
- 27. Abdulhussain, S.H.; Mahmmod, B.M.; Naser, M.A.; Alsabah, M.Q.; Ali, R.; Al-Haddad, S. A robust handwritten numeral recognition using hybrid orthogonal polynomials and moments. *Sensors* **2021**, *21*, 1999. [CrossRef] [PubMed]
- 28. Padierna, L.C.; Carpio, M.; Rojas-Domínguez, A.; Puga, H.; Fraire, H. A novel formulation of orthogonal polynomial kernel functions for SVM classifiers: The Gegenbauer family. *Pattern Recognit.* **2018**, *84*, 211–225. [CrossRef]
- 29. Fischer, B. Polynomial Based Iteration Methods for Symmetric Linear Systems; SIAM: Philadelphia, PA, USA, 2011.

- Pedregosa, F.; Scieur, D. Acceleration through spectral density estimation. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 13-18 July 2020; pp. 7553–7562.
- Kuriakose, M.; Chen, W.; He, Z.; Sikora, A.; Zhang, P.; Zhang, Z.; Qiu, W.; Hsu, D.; McMunn-Coffran, C.; Brown, S.; et al. Expression Data from Head and Neck Squamous Cell Carcinoma. 2007. Available online: https://www.ncbi.nlm.nih.gov/geo/ query/acc.cgi?acc=GSE6631 (accessed on 7 March 2022).
- National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce. ETOPO1, Global 1 Arc-minute Ocean Depth and Land Elevation from the US National Geophysical Data Center (NGDC). 2011. Available online: https://rda.ucar.edu/ datasets/ds759.4/ (accessed on 7 March 2022).