ASYMPTOTIC REDUNDANCIES FOR UNIVERSAL QUANTUM CODING

CHRISTIAN KRATTENTHALER AND PAUL B. SLATER

Abstract. Clarke and Barron have recently shown that the Jeffreys' invariant prior of Bayesian theory yields the common asymptotic (minimax and maximin) redundancy of universal data compression in a parametric setting. We seek a possible analogue of this result for the two-level *quantum* systems. We restrict our considerations to prior probability distributions belonging to a certain one-parameter family, $q(u)$, $-\infty < u < 1$. Within this setting, we are able to compute exact redundancy formulas, for which we find the asymptotic limits. We compare our quantum asymptotic redundancy formulas to those derived by naively applying the classical counterparts of Clarke and Barron, and find certain common features. Our results are based on formulas we obtain for the eigenvalues and eigenvectors of $2^n \times 2^n$ (Bayesian density) matrices, $\zeta_n(u)$. These matrices are the weighted averages (with respect to $q(u)$) of all possible tensor products of n identical 2×2 density matrices, representing the two-level quantum systems. We propose a form of *universal* coding for the situation in which the density matrix describing an ensemble of quantum signal states is unknown. A sequence of n signals would be projected onto the dominant eigenspaces of $\zeta_n(u)$.

1. Introduction

A theorem has recently been proven [30, 47] (cf. [7, 19, 35]), in the context of quantum information theory [7, 40], that is analogous to the noiseless coding theorem of classical information theory. In the quantum result, the von Neumann entropy [39, 58],

$$S(\rho) = -\operatorname{Tr} \rho \log \rho \tag{1.1}$$

(equalling the Shannon entropy of the probability distribution formed by the eigenvalues of ρ) of the density matrix,

$$\rho = \sum_a p(a)\pi_a, \tag{1.2}$$

Key words and phrases. Quantum information theory, two-level quantum systems, universal data compression, asymptotic redundancy, Jeffreys' prior, Bayes redundancy, Schumacher compression, ballot paths, Dyck paths, relative entropy, Bayesian density matrices, quantum coding, Bayes codes, monotone metric, symmetric logarithmic derivative, Kubo-Mori/Bogoliubov metric.

Krattenthaler's research was supported in part by MSRI, through NSF grant DMS-9022140.

describing an ensemble of *pure* quantum signal states, is equal to $\log 2 \approx .693147$ times the number of quantum bits ("qubits") — that is, the number of two-dimensional Hilbert spaces — necessary to represent the signal faithfully. (Although the binary logarithm is usually used in the quantum coding literature, we employ the natural logarithm throughout this paper, chiefly to facilitate comparisons of our results with those of Clarke and Barron [16, 17, 18]. $p(a)$ is the probability of the message a from a particular source coded into a "signal state" — having a state vector denoted by the ket $|a_M\rangle$ — of a quantum system M. The density matrices π_a are the projections $\pi_a = |a_M\rangle\langle a_M|$, with $\langle a_M|$ being a bra in the dual Hilbert space.)

The proof of the quantum coding theorem is based on the existence of a "typical subspace" Λ of the 2^n-dimensional Hilbert space of n qubits, which has the property that, with high probability, a sample of n qubits has almost unit projection onto Λ. Since it has been shown that the dimension of Λ is $e^{nS(\rho)}$, the operation that the data compressor (a unitary transformation mapping n-qubit strings to n-qubit strings) should perform involves "transposing" the subspace Λ into the Hilbert space of a smaller block of $nS(\rho)/.693147$ qubits [19]. (Lo [35] has generalized this work for an ensemble of *mixed* quantum signal states.)

In this study we dispense with the assumption that *a priori* information (other than its dimensionality) is available regarding ρ. Somewhat similarly motivated, Calderbank and Shor [12] modified the definition of *fidelity* — a measure of the success of transmission of quantum states — because "previous papers discuss channels that transmit some distribution of states given *a priori*, whereas we want our channel to faithfully transmit any pure input state". They took as their measure, the fidelity for the pure state transmitted *least* faithfully.

Proceeding in a noninformative Bayesian framework [9, 49, 50, 51], we seek to extend to the two-level quantum systems, recent results of Clarke and Barron [16, 17, 18] giving various forms of the asymptotic redundancy of universal data compression for parameterized families of probability distributions. "The redundancy is the excess of the [coding] cost over the entropy. The goal of data compression is to diminish redundancy" ([33], reviewed in [20]). "The idea of universal coding, suggested by Kolmogorov, is to construct a code for data sequences such that asymptotically, as the length of the sequence increases, the mean per symbol code length would approach the entropy of whatever process in a family has generated the data" [45]. For an extensive commentary on the results of Clarke and Barron, see [45]. Also see [15], for some recent related research, as well as a discussion of various rationales that have been employed for using the (classical) Jeffreys' prior — a possible quantum counterpart of which will be of interest here — for Bayesian purposes, cf. [32]. Let us also bring to the attention of the reader that in a brief review of [17], the noted statistician, I. J. Good, commented that Clarke and Barron " have presumably overlooked the reviewer's work" and cited, in this regard [27, 28]. (It should be noted that in these papers, Good uses a more general objective function — a two-parameter utility —

than the relative entropy, chosen by Clarke and Barron over alternative measures [16, p. 454]. Good does conclude that Jeffreys' invariant prior is the minimax, that is, the least favorable, prior when the utility is the "weight of the evidence" in the sense of C. S. Pierce, that is, the relative entropy.)

Clarke and Barron [16, 17, 18] found the asymptotic redundancy to be given by

$$\frac{d}{2}\log\frac{n}{2\pi e}+\frac{1}{2}\log\det I(\theta)-\log w(\theta)+o(1). \tag{1.3}$$

Here, θ is a d-dimensional vector of variables parameterizing a family (manifold) of probability distributions. $I(\theta)$ is the $d\times d$ Fisher information matrix — the negative of the expected value of the Hessian of the logarithm of the density function — and $w(\theta)$ is the prior density. The asymptotic *minimax* redundancy was shown to be [17, 18]

$$\frac{d}{2}\log\frac{n}{2\pi e}+\log\int_K\sqrt{\det I(\theta)}\,d\theta+o(1), \tag{1.4}$$

where K is a compact set in the interior of the domain of the parameters.

In this investigation, instead of probability densities as in [16, 17, 18], we employ density matrices (nonnegative definite Hermitian matrices of unit trace) and instead of the classical form of the relative entropy (the Kullback–Leibler information measure), its quantum counterpart [39, 58] (cf. [44]),

$$S(\rho_1,\rho_2)=\mathrm{Tr}\,\rho_1(\log\rho_1-\log\rho_2), \tag{1.5}$$

that is, the relative entropy of the density matrix ρ_1 with respect to ρ_2.

The three-dimensional convex set of 2×2 density matrices that will be the focus of our study has members representable in the form,

$$\rho=\frac{1}{2}\begin{pmatrix}1+z & x-iy\\ x+iy & 1-z\end{pmatrix}\quad . \tag{1.6}$$

Such matrices correspond, in a one-to-one fashion, to the standard (complex) two-level quantum systems — notably, those of spin-1/2 (electrons, protons, ...) and massless spin-1 particles (photons). (If we set $x=y=0$ in (1.6), we recover a classical binomial distribution, with the probability of "success", say, being $(1+z)/2$ and of "failure", $(1-z)/2$. Setting either x or y to zero, puts us in the framework of real — as opposed to complex — quantum mechanics.) The points (x,y,z) must lie within the unit ball ("Bloch sphere" [11]), $x^2+y^2+z^2\leq 1$, due to the requirement for ρ of nonnegative eigenvalues. (The points on the bounding spherical surface, $x^2+y^2+z^2=1$, corresponding to the pure states, will be shown to exhibit nongeneric behavior, see (2.38) and the respective comments in Sec. 3 (cf. [24]).) We have, for (1.6), using spherical coordinates (r,ϑ,ϕ), so that $r=(x^2+y^2+z^2)^{1/2}$,

$$S(\rho)=-\frac{(1-r)}{2}\log\frac{(1-r)}{2}-\frac{(1+r)}{2}\log\frac{(1+r)}{2}. \tag{1.7}$$

A composite system of n identical independent (unentangled) two-level quantum systems is represented by the $2^n \times 2^n$ density matrix $\overset{n}{\otimes}\rho$ — possessing a von Neumann entropy $nS(\rho)$ [39, 58]. (In noncommutative probability theory, independence can be based on free products instead of tensor products [55]. Along with the real and complex forms of quantum mechanics, a quaternionic version exists [22], for which the [presumed] quantum Jeffreys' prior has been found for the two-level systems — corresponding to the *five*-dimensional unit ball/"Bloch sphere" [49]. However, the definition of a tensor product is somewhat problematical in this context [1, 21].)

In [49] it was argued that the quantum Fisher information matrix (requiring — due to noncommutativity — the computation of symmetric logarithmic derivatives [42]) for the density matrices (1.6) should be taken to be of the form

$$I(\theta) = \frac{1}{(1-x^2-y^2-z^2)} \begin{pmatrix} 1-y^2-z^2 & xy & xz \\ xy & 1-x^2-z^2 & yz \\ xz & yz & 1-x^2-y^2 \end{pmatrix} . \tag{1.8}$$

The quantum counterpart of the Jeffreys' prior was, then, taken to be the normalized form (dividing by π^2) of the square root of the determinant of (1.8), that is,

$$(1-x^2-y^2-z^2)^{-1/2}/\pi^2. \tag{1.9}$$

Analogously, the *classical* Jeffreys' prior is proportional to the square root of the determinant of the *classical* Fisher information matrix [9].

On the basis of the result of Clarke and Barron [17, 18] that the Jeffreys' prior yields the asymptotic common (minimax and maximin) redundancy (that is, the least favorable and reference priors are the same), it was conjectured [52] that its assumed quantum counterpart (1.9) would have similar properties, as well. (The Jeffreys' prior has been "shown to be a minimax solution in a — two person — zero sum game, where the statistician chooses the 'non-informative' prior and nature chooses the 'true' prior" [9, 31]. Quantum mechanics itself has been asserted to arise from a Fisher-information transfer zero sum game [23].) To examine this possibility, (1.9) was embedded as a specific member ($u = .5$) of a one-parameter family of spherically-symmetric/unitarily-invariant probability densities,

$$q(u) = \frac{\Gamma(5/2-u)}{\pi^{3/2}\,\Gamma(1-u)\,(1-x^2-y^2-z^2)^u}, \quad -\infty < u < 1. \tag{1.10}$$

(Under unitary transformations of ρ, the assigned probability is invariant.) For $u = 0$, we obtain a uniform distribution over the unit ball. (This has been used as a prior over the two-level quantum systems, at least, in one study [34].) For $u \to 1$, the uniform distribution over the spherical boundary (the locus of the pure states) is approached. (This is often employed as a prior, for example [29, 34, 36].) For $u \to -\infty$, a Dirac distribution concentrated at the origin (corresponding to the fully mixed state) is approached.

Embeddings of (1.9) in other (possibly, multiparameter) families are, of course, possible and may be pursued in further research. Ideally, we would aspire to formally demonstrate — if it is, in fact, so — that (1.9) can be uniquely characterized *vis-à-vis* *all* other possible probability distributions over the unit ball. Due to the present lack of any such fully rigorous treatment, analogous to that of Clarke and Barron, we rely upon an exploratory heuristic computational strategy. This involves averaging $\overset{n}{\otimes}\rho$ with respect to $q(u)$. Doing so yields a one-parameter family of $2^n \times 2^n$ Bayesian density matrices (Bayes codes or estimators [18, 16, 37]), $\zeta_n(u)$, $-\infty < u < 1$, exhibiting highly interesting properties.

We explicitly find (in Sec. 2) the eigenvalues and eigenvectors of the matrices $\zeta_n(u)$ and determine the relative entropy (1.5) of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$. We do this by using identities for hypergeometric series and some combinatorics. (It is also possible to obtain some of our results by making use of representation theory of $SU(2)$. An even more general result was derived by combining these two approaches. We comment on this issue at the end of Sec. 3.)

The matrices $\zeta_n(u)$ should prove useful for the *universal* version of Schumacher data compression [7, 19, 30, 47] by projecting blocks of n signals (qubits) onto those "typical" subspaces of 2^n-dimensional Hilbert space corresponding to as many of the dominant eigenvalues of $\zeta_n(u)$ as it takes to exceed a sum $1-\epsilon$. (This can be accomplished by a unitary transformation, the inverse of which would be used in the decoding step [7]. In the corresponding nonuniversal quantum coding context, the projection onto the dominant eigenvalues of $\overset{n}{\otimes}\rho$ yields fidelity greater than $1-2\epsilon$ [30] and distortion less than 2ϵ [35], cf. [5].) For all u, the leading one of the $\lfloor \frac{n}{2} \rfloor + 1$ distinct eigenvalues has multiplicity $n+1$, and belongs to the $(n+1)$-dimensional (Bose–Einstein) symmetric subspace [3]. (Projection onto the symmetric subspace has been proposed as a method for stabilizing quantum computations, including quantum state storage [4].) For $u = 1/2$, the leading eigenvalue can be obtained by dividing the $n+1$-st Catalan number — that is, $\frac{1}{n+2}\binom{2(n+1)}{n+1}$ — by 4^n. (The Catalan numbers "are probably the most frequently occurring combinatorial numbers after the binomial coefficients" [53].)

Let us (naively) attempt to apply the formulas of Clarke and Barron [17, 18] — (1.4) and (1.3) above — to the quantum context under investigation here. We do this by setting d to 3 (the dimensionality of the unit ball — which we take as K), $\det I(\theta)$ to $(1-x^2-y^2-z^2)^{-1}$ (cf. (1.8)), so that $\int_K \sqrt{\det I(\theta)}\, d\theta$ is π^2, and $w(\theta)$ to $q(u)$. Then, we obtain from the expression for the asymptotic *minimax* redundancy (1.4),

$$\frac{3}{2}(\log n - \log 2 - 1) + \frac{1}{2}\log \pi + o(1), \tag{1.11}$$

and from the expression for the asymptotic redundancy itself (1.3),

$$\frac{3}{2}(\log n - \log 2 - 1) - (1-u)\log(1-r^2) + \log\Gamma(1-u) - \log\Gamma\left(\frac{5}{2}-u\right) + o(1) \tag{1.12}$$

We shall (in Sec. 3) compare these two formulas, (1.11) and (1.12), with the results of Sec. 2 and find some striking similarities and coincidences, particularly associated with the fully mixed state ($r=0$). These findings will help to support the working hypothesis of this study — that there are meaningful extensions to the quantum domain of the (commutative probabilistic) theorems of Clarke and Barron. However, we find that although the minimax property of the Jeffreys' prior appears to carry over, the maximin property does not strictly, but only in an approximate sense. In any case, we can not formally rule out the possibility that the actual global (perhaps common) minimax and maximin are achieved for probability distributions not belonging to the one-parameter family $q(u)$.

Let us point out to the reader the quite recent important work of Petz and Sudar [42]. They demonstrated that in the quantum case — in contrast to the classical situation in which there is, as originally shown by Chentsov [14], essentially only one monotone metric and, therefore, essentially only one form of the Fisher information — there exists an infinitude of such metrics. "The monotonicity of the Riemannian metric g is crucial when one likes to imitate the geometrical approach of [Chentsov]. An infinitesimal statistical distance has to be monotone under stochastic mappings. We note that the monotonicity of g is a strengthening of the concavity of the von Neumann entropy. Indeed, positive definiteness of g is equivalent to the strict concavity of the von Neumann entropy ... and monotonicity is much more than positivity" [41].

The monotone metrics on the space of density matrices are given [42] by the operator monotone functions $f(t):\mathbb{R}^+\to\mathbb{R}^+$, such that $f(1)=1$ and $f(t)=tf(1/t)$. For the choice $f=(1+t)/2$, one obtains the minimal metric (of the symmetric logarithmic derivative), which serves as the basis of our analysis here. "In accordance with the work of Braunstein and Caves, this seems to be the canonical metric of parameter estimation theory. However, expectation values of certain relevant observables are known to lead to statistical inference theory provided by the maximum entropy principle or the minimum relative entropy principle when *a priori* information on the state is available. The best prediction is a kind of generalized Gibbs state. On the manifold of those states, the differentiation of the entropy functional yields the Kubo-Mori/Bogoliubov metric, which is different from the metric of the symmetric logarithmic derivative. Therefore, more than one privileged metric shows up in quantum mechanics. The exact clarification of this point requires and is worth further studies" [42]. It remains a possibility, then, that a monotone metric other than the minimal one (which corresponds to $q(.5)$, that is (1.9)) may yield a common global asymptotic minimax and maximin redundancy, thus, fully paralleling the classical/nonquantum results of Clarke and Barron [16, 17, 18]. We intend to investigate such a possibility, in particular, for the Kubo-Mori/Bogoliubov metric [41, 42, 43].

2. Analysis of a One-Parameter Family of Bayesian Density Matrices

In this section, we implement the analytical approach described in the Introduction to extending the work of Clarke and Barron [17, 18] to the realm of quantum mechanics, specifically, the two-level systems. Such systems are representable by density matrices ρ of the form (1.6). A composite system of n independent (unentangled) and identical two-level quantum systems is, then, represented by the n-fold tensor product $\overset{n}{\otimes}\rho$. In Theorem 1 of Sec. 2.1, we average $\overset{n}{\otimes}\rho$ with respect to the one-parameter family of probability densities $q(u)$ defined in (1.10), obtaining the Bayesian density matrices $\zeta_n(u)$ and formulas for their 2^{2n} entries. Then, in Theorem 2 of Sec. 2.2, we are able to explicitly determine the 2^n eigenvalues and eigenvectors of $\zeta_n(u)$. Using these results, in Sec. 2.3, we compute the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$. Then, in Sec. 2.4, we obtain the asymptotics of this relative entropy for $n \to \infty$. In Sec. 2.5, we compute the asymptotics of the von Neumann entropy (see (1.1)) of $\zeta_n(u)$. All these results will enable us, in Sec. 3, to ascertain to what extent the results of Clarke and Barron could be said to carry over to the quantum domain.

2.1. Entries of the Bayesian density matrices $\zeta_n(u)$.

The n-fold tensor product $\overset{n}{\otimes}\rho$ is a $2^n \times 2^n$ matrix. To refer to specific rows and columns of $\overset{n}{\otimes}\rho$, we index them by subsets of the n-element set $\{1, 2, \ldots, n\}$. We choose to employ this notation instead of the more familiar use of binary strings, in order to have a more succinct way of writing our formulas. For convenience, we will subsequently write $[n]$ for $\{1, 2, \ldots, n\}$. Thus, $\overset{n}{\otimes}\rho$ can be written in the form

$$\overset{n}{\otimes}\rho = \left(R_{IJ}\right)_{I,J\in[n]},$$

where

$$R_{IJ} = \frac{1}{2^n}(1+z)^{n_{\in\in}}(1-z)^{n_{\notin\notin}}(x+iy)^{n_{\notin\in}}(x-iy)^{n_{\in\notin}}, \tag{2.1}$$

with $n_{\in\in}$ denoting the number of elements of $[n]$ contained in both I and J, $n_{\notin\notin}$ denoting the number of elements *not* in both I and J, $n_{\notin\in}$ denoting the number of elements not in I but in J, and $n_{\in\notin}$ denoting the number of elements in I but not in

J. In symbols,

$$\begin{aligned} n_{\in\in} &= |I \cap J|, \\ n_{\notin\notin} &= |[n]\backslash(I \cup J)|, \\ n_{\notin\in} &= |J\backslash I|, \\ n_{\in\notin} &= |I\backslash J|. \end{aligned}$$

We consider the average $\zeta_n(u)$ of $\overset{n}{\otimes}\rho$ with respect to the probability density $q(u)$ defined in (1.10) taken over the unit sphere $\{(x, y, z) : x^2 + y^2 + z^2 \le 1\}$. This average can be described explicitly as follows.

Theorem 1. *The average $\zeta_n(u)$,*

$$\int_{x^2+y^2+z^2\le 1} \big(\overset{n}{\otimes}\rho\big)\, q(u)\, dx\, dy\, dz,$$

equals the matrix $(Z_{IJ})_{I,J\in[n]}$, where

$$\begin{aligned} Z_{IJ} = \delta_{n_{\notin\in},n_{\in\notin}} &\left(\tfrac{n-n_{\in\in}-n_{\notin\notin}}{2}\right)! \\ &\times \frac{1}{2^n} \frac{\Gamma\left(\frac{5}{2}-u\right)\,\Gamma\left(2+\frac{n}{2}+\frac{n_{\in\in}}{2}-\frac{n_{\notin\notin}}{2}-u\right)\,\Gamma\left(2+\frac{n}{2}+\frac{n_{\notin\notin}}{2}-\frac{n_{\in\in}}{2}-u\right)}{\Gamma\left(\frac{5}{2}+\frac{n}{2}-u\right)\,\Gamma\left(2+\frac{n}{2}-u\right)\,\Gamma\left(2+\frac{n}{2}-\frac{n_{\in\in}}{2}-\frac{n_{\notin\notin}}{2}-u\right)}. \end{aligned} \tag{2.2}$$

Here, $\delta_{i,j}$ denotes the Kronecker delta, $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise.

Remark. It is important for later considerations to observe that because of the term $\delta_{n_{\notin\in},n_{\in\notin}}$ in (2.2) the entry Z_{IJ} is nonzero if and only if the sets I and J have the same cardinality. If I and J have the same cardinality, c say, then Z_{IJ} only depends on $n_{\in\in}$, the number of common elements of I and J, since in this case $n_{\notin\notin}$ is expressible as $n - 2c + n_{\in\in}$.

Proof of Theorem 1. To compute Z_{IJ}, we have to compute the integral

$$\int_{x^2+y^2+z^2\le 1} R_{IJ}\, q(u)\, dx\, dy\, dz. \tag{2.3}$$

For convenience, we treat the case that $n_{\in\in} \ge n_{\notin\notin}$ and $n_{\notin\in} \ge n_{\in\notin}$. The other four cases are treated similarly.

First, we rewrite the matrix entries R_{IJ},

$$\begin{aligned}
&\frac{1}{2^n}(1+z)^{n_{\in\in}}(1-z)^{n_{\notin\notin}}(x+iy)^{n_{\notin\in}}(x-iy)^{n_{\in\notin}}\\
&\qquad=\frac{1}{2^n}(1-z^2)^{n_{\in\in}}(1-z)^{n_{\notin\notin}-n_{\in\in}}(x^2+y^2)^{n_{\notin\in}}(x-iy)^{n_{\in\notin}-n_{\notin\in}}\\
&\qquad=\frac{1}{2^n}\sum_{j,k,l\geq 0}(-1)^{j+k}(-i)^l\binom{n_{\in\in}}{j}\binom{n_{\notin\notin}-n_{\in\in}}{k}\binom{n_{\in\notin}-n_{\notin\in}}{l}\\
&\qquad\qquad\cdot z^{2j+k}(x^2+y^2)^{n_{\notin\in}}x^{n_{\in\notin}-n_{\notin\in}-l}y^l. \qquad (2.4)
\end{aligned}$$

Of course, in order to compute the integral (2.3), we transform the Cartesian coordinates into polar coordinates,

$$\begin{aligned}
x &= r\sin\vartheta\cos\varphi\\
y &= r\sin\vartheta\sin\varphi\\
z &= r\cos\vartheta,\\
0\leq\varphi\leq 2\pi,&\ \ 0\leq\vartheta\leq\pi.
\end{aligned}$$

Thus, using (2.4), the integral (2.3) is transformed into

$$\begin{aligned}
&\frac{1}{2^n}\sum_{j,k,l\geq 0}\int_0^1\int_0^\pi\int_0^{2\pi}(-1)^{j+k}(-i)^l\binom{n_{\in\in}}{j}\binom{n_{\notin\notin}-n_{\in\in}}{k}\binom{n_{\in\notin}-n_{\notin\in}}{l}\\
&\qquad\cdot r^{2j+k+n_{\notin\in}+n_{\in\notin}+2}\left(\cos^{2j+k}\vartheta\right)\left(\sin^{n_{\notin\in}+n_{\in\notin}+1}\vartheta\right)\\
&\qquad\cdot\left(\cos^{n_{\in\notin}-n_{\notin\in}-l}\varphi\right)\left(\sin^l\varphi\right)\frac{\Gamma(5/2-u)}{\pi^{3/2}\,\Gamma(1-u)\,(1-r^2)^u}\,d\varphi\,d\vartheta\,dr. \qquad (2.5)
\end{aligned}$$

To evaluate this triple integral we use the following standard formulas:

$$\int_0^\pi \sin^{2M}\vartheta\,\cos^{2N}\vartheta\,d\vartheta=\pi\frac{(2M-1)!!\,(2N-1)!!}{(2M+2N)!!}, \qquad (2.6a)$$

$$\int_0^\pi \sin^{2M+1}\vartheta\,\cos^{2N}\vartheta\,d\vartheta=2\frac{(2M)!!\,(2N-1)!!}{(2M+2N+1)!!} \qquad (2.6b)$$

$$\text{and}\quad \int_0^{2\pi} \sin^{2M+1}\vartheta\,\cos^{2N}\vartheta\,d\vartheta=0, \qquad (2.6c)$$

$$\int_0^\pi \sin^{2M}\vartheta\,\cos^{2N+1}\vartheta\,d\vartheta=0, \qquad (2.6d)$$

$$\int_0^\pi \sin^{2M+1}\vartheta\,\cos^{2N+1}\vartheta\,d\vartheta=0, \qquad (2.6e)$$

for any nonnegative integers M and N. Furthermore, we need the beta integral

$$\int_0^1 \frac{r^m}{(1-r^2)^u}\,dr = \frac{\Gamma\left(\frac{m+1}{2}\right)\,\Gamma(1-u)}{2\,\Gamma\left(\frac{m+3}{2}-u\right)}. \tag{2.7}$$

Now we consider the integral over φ in (2.5). Using (2.6c) and (2.6d), we see that each summand in (2.5) vanishes if $n_{\notin\in}$ has a parity different from $n_{\in\notin}$. On the other hand, if $n_{\notin\in}$ has the same parity as $n_{\in\notin}$, then we can evaluate the integrals over φ using (2.6a) and (2.6e). Discarding for a moment the terms independent of φ and l, we have

$$\begin{aligned}
&\sum_{l\ge0}\int_0^{2\pi}(-i)^l\binom{n_{\in\notin}-n_{\notin\in}}{l}\left(\cos^{n_{\in\notin}-n_{\notin\in}-l}\varphi\right)\left(\sin^l\varphi\right)\,d\varphi\\
&\quad=\sum_{l\ge0}(-1)^l\binom{n_{\in\notin}-n_{\notin\in}}{2l}2\pi\,\frac{(2l-1)!!\,(n_{\in\notin}-n_{\notin\in}-2l-1)!!}{(n_{\in\notin}-n_{\notin\in})!!}\\
&\quad=2\pi\frac{(n_{\in\notin}-n_{\notin\in}-1)!!}{(n_{\in\notin}-n_{\notin\in})!!}\sum_{l\ge0}\binom{(n_{\in\notin}-n_{\notin\in})/2}{l}(-1)^l\\
&\quad=2\pi\,\delta_{n_{\in\notin},n_{\notin\in}},
\end{aligned}$$

the last line being due to the binomial theorem. These considerations reduce (2.5) to

$$\begin{aligned}
\delta_{n_{\in\notin},n_{\notin\in}}\frac{1}{2^n}\sum_{j,k\ge0}\int_0^1\int_0^\pi(-1)^{j+k}\binom{n_{\in\in}}{j}\binom{n_{\notin\notin}-n_{\in\in}}{k}\\
\cdot r^{2j+k+2n_{\notin\in}+2}\left(\cos^{2j+k}\vartheta\right)\left(\sin^{2n_{\notin\in}+1}\vartheta\right)\frac{2\,\Gamma(5/2-u)}{\pi^{1/2}\,\Gamma(1-u)\,(1-r^2)^u}\,d\vartheta\,dr.
\end{aligned}$$

Using (2.6c), (2.6e) and (2.7) this can be further simplified to

$$\begin{aligned}
\delta_{n_{\in\notin},n_{\notin\in}}\frac{1}{2^n}\sum_{j,k\ge0}(-1)^j\binom{n_{\in\in}}{j}\binom{n_{\notin\notin}-n_{\in\in}}{2k}\frac{2\,(2j+2k-1)!!\,(2n_{\notin\in})!!}{(2j+2k+2n_{\notin\in}+1)!!}\\
\cdot\frac{\Gamma(j+k+n_{\notin\in}+3/2)\,\Gamma(1-u)}{2\,\Gamma(j+k+n_{\notin\in}+5/2-u)}\,\frac{2\,\Gamma(5/2-u)}{\pi^{1/2}\,\Gamma(1-u)}.
\end{aligned}\tag{2.8}$$

Next we interchange sums over j and k and write the sum over k in terms of the standard hypergeometric notation

$${}_rF_s\left[\begin{matrix}a_1,\dots,a_r\\ b_1,\dots,b_s\end{matrix};z\right]=\sum_{k=0}^\infty\frac{(a_1)_k\cdots(a_r)_k}{k!\,(b_1)_k\cdots(b_s)_k}z^k\ ,$$

where the shifted factorial $(a)_k$ is given by $(a)_k := a(a+1)\cdots(a+k-1)$, $k \ge 1$, $(a)_0 := 1$. Thus we can write (2.8) in the form

$$\delta_{n_{\in\notin},n_{\notin\in}} \frac{1}{2^n} \sum_{k\ge 0} \binom{n_{\notin\notin}-n_{\in\in}}{2k} \frac{(2k-1)!!\, n_{\notin\in}!\, \Gamma\left(\frac{5}{2}-u\right)}{2^{k+1}\, \Gamma\left(\frac{5}{2}+k+n_{\notin\in}-u\right)} \cdot {}_2F_1\left[\begin{matrix}\frac{1}{2}+k, -n_{\in\in}\\ \frac{5}{2}+k+n_{\notin\in}-u\end{matrix}; 1\right]. \tag{2.9}$$

The ${}_2F_1$ series can be summed by means of Gauß' ${}_2F_1$ summation (see e.g. [48, (1.7.6); Appendix (III.3)])

$${}_2F_1\left[\begin{matrix}a, b\\ c\end{matrix}; 1\right] = \frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)}, \tag{2.10}$$

provided the series terminates or $\mathrm{Re}(c-a-b) \ge 0$. Applying (2.10) to the ${}_2F_1$ in (2.9) (observe that it is terminating) and writing the sum over k as a hypergeometric series, the expression (2.9) becomes

$$\delta_{n_{\in\notin},n_{\notin\in}} \frac{1}{2^n} \frac{\Gamma(2+n_{\in\in}+n_{\notin\in}-u)\,\Gamma\left(\frac{5}{2}-u\right)\, n_{\notin\in}!}{\Gamma\left(\frac{5}{2}+n_{\in\in}+n_{\notin\in}-u\right)\, \Gamma(2+n_{\notin\in}-u)} \times {}_2F_1\left[\begin{matrix}\frac{n_{\in\in}}{2}-\frac{n_{\notin\notin}}{2}, \frac{1}{2}+\frac{n_{\in\in}}{2}-\frac{n_{\notin\notin}}{2}\\ \frac{5}{2}+n_{\in\in}+n_{\notin\in}-u\end{matrix}; 1\right].$$

Another application of (2.10) gives

$$\delta_{n_{\in\notin},n_{\notin\in}} \frac{1}{2^n} \times \frac{\Gamma(2+n_{\in\in}+n_{\notin\in}-u)\,\Gamma(2+n_{\notin\notin}+n_{\notin\in}-u)\,\Gamma\left(\frac{5}{2}-u\right)\, n_{\notin\in}!}{\Gamma\left(\frac{5}{2}+\frac{n_{\in\in}}{2}+\frac{n_{\notin\notin}}{2}+n_{\notin\in}-u\right)\, \Gamma\left(2+\frac{n_{\in\in}}{2}+\frac{n_{\notin\notin}}{2}+n_{\notin\in}-u\right)\, \Gamma(2+n_{\notin\in}-u)}. \tag{2.11}$$

Trivially, we have $n = n_{\in\in}+n_{\notin\notin}+n_{\notin\in}+n_{\in\notin}$. Since (2.11) vanishes unless $n_{\notin\in} = n_{\in\notin}$, we can substitute $(n-n_{\in\in}-n_{\notin\notin})/2$ for $n_{\notin\in}$ in the arguments of the gamma functions. Thus, we see that (2.11) equals (2.2). This completes the proof of the Theorem. ☐

2.2. **Eigenvalues and eigenvectors of the Bayesian density matrices $\zeta_n(u)$.** With the explicit description of the result $\zeta_n(u)$ of averaging $\overset{n}{\otimes}\rho$ with respect to $q(u)$ at our disposal, we now proceed to describe the eigenvalues and eigenspaces of $\zeta_n(u)$. The eigenvalues are given in Theorem 2. Lemma 4 gives a complete set of eigenvectors of $\zeta_n(u)$. The reader should note that, though complete, this is simply a set of linearly independent eigenvectors and not a fully orthogonal set.

Theorem 2. *The eigenvalues of the $2^n \times 2^n$ matrix $\zeta_n(u)$, the entries of which are given by* (2.2)*, are*

$$\lambda_d = \frac{1}{2^n} \frac{\Gamma\left(\frac{5}{2} - u\right) \Gamma(2+n-d-u)\,\Gamma(1+d-u)}{\Gamma\left(\frac{5}{2} + \frac{n}{2} - u\right) \Gamma(2+\frac{n}{2}-u)\,\Gamma(1-u)}, \quad d = 0, 1, \dots, \left\lfloor \frac{n}{2} \right\rfloor, \tag{2.12}$$

with respective multiplicities

$$\frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d}. \tag{2.13}$$

The Theorem will follow from a sequence of Lemmas. We state the Lemmas first, then prove Theorem 2 assuming the truth of the Lemmas, and after that provide proofs of the Lemmas.

In the first Lemma some eigenvectors of the matrix $\zeta_n(u)$ are described. Clearly, since $\zeta_n(u)$ is a $2^n \times 2^n$ matrix, the eigenvectors are in 2^n-dimensional space. As we did previously, we index coordinates by subsets of $[n]$, so that a generic vector is $(x_S)_{S\in[n]}$. In particular, given a subset T of $[n]$, the symbol e_T denotes the standard unit vector with a 1 in the T-th coordinate and 0 elsewhere, i.e., $e_T = (\delta_{S,T})_{S\in[n]}$.

Now let d, s be integers with $0 \le d \le s \le n-d$ and let A and B be two disjoint d-element subsets A and B of $[n]$. Then we define the vector $v_{d,s}(A,B)$ by

$$v_{d,s}(A,B) := \sum_{\substack{X\subseteq A\\ Y\subseteq[n]\setminus(A\cup B),\ |Y|=s-d}} (-1)^{|X|}\, e_{X\cup X'\cup Y}, \tag{2.14}$$

where X' is the "*complement of* X *in* B" by which we mean that if X consists of the i_1-, i_2-, … -largest elements of A, $i_1 < i_2 < \cdots$, then X' consists of all elements of B *except for* the i_1-, i_2-, … -largest elements of B. For example, let $n = 7$. Then the vector $v_{2,3}(\{1,3\},\{2,5\})$ is given by

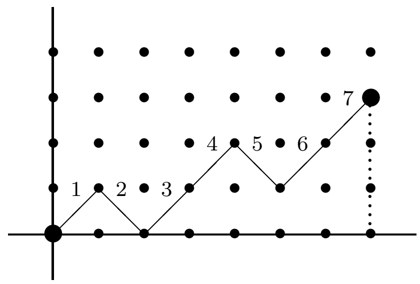
$$e_{\{2,4,5\}} + e_{\{2,5,6\}} + e_{\{2,5,7\}} - e_{\{1,4,5\}} - e_{\{1,5,6\}} - e_{\{1,5,7\}} \\ - e_{\{2,3,4\}} - e_{\{2,3,6\}} - e_{\{2,3,7\}} + e_{\{1,3,4\}} + e_{\{1,3,6\}} + e_{\{1,3,7\}}. \tag{2.15}$$

(In this special case, the possible subsets X of $A = \{1,3\}$ in the sum in (2.14) are $\emptyset$, $\{1\}$, $\{3\}$, $\{1,3\}$, with corresponding complements in $B = \{2,5\}$ being $\{2,5\}$, $\{5\}$, $\{2\}$, $\emptyset$, respectively, and the possible sets Y are $\{4\}$, $\{6\}$, $\{7\}$.) Observe that all sets $X \cup X' \cup Y$ which occur as indices in (2.14) have the same cardinality s.

Lemma 3. *Let d, s be integers with $0 \le d \le s \le n-d$ and let A and B be disjoint d-element subsets of $[n]$. Then $v_{d,s}(A,B)$ as defined in* (2.14) *is an eigenvector of the matrix $\zeta_n(u)$, the entries of which are given by* (2.2)*, for the eigenvalue λ_d, where λ_d is given by* (2.12).

We want to show that the multiplicity of λ_d equals the expression in (2.13). Of course, Lemma 3 gives many more eigenvectors for λ_d. Therefore, in order to describe a basis for the corresponding eigenspace, we have to restrict the collection of vectors in Lemma 3.

We do this in the following way. Fix d, $0 \le d \le \lfloor n/2 \rfloor$. Let P be a lattice path in the plane integer lattice $\mathbb{Z}^2$, starting in $(0,0)$, consisting of $n-d$ up-steps $(1,1)$ and d down-steps $(1,-1)$, which never goes below the x-axis. Figure 1 displays an example with $n=7$ and $d=2$. Clearly, the end point of P is $(n, n-2d)$. We call a lattice path which starts in $(0,0)$ and never goes below the x-axes a *ballot path*. (This terminology is motivated by its relation to the (two-candidate) *ballot problem*, see e.g. [38, Ch. 1, Sec. 1]. An alternative term for ballot path which is often used is "Dyck path", see e.g. [56, p. I-12].) We will use the abbreviation "b.p." for "ballot path" in displayed formulas.



Ballot paths

Figure 1

Given such a lattice path P, label the steps from 1 to n, as is indicated in Figure 1. Then define A_P to be set of all labels corresponding to the first d up-steps of P and B_P to be set of all labels corresponding to the d down-steps of P. In the example of Figure 1 we have for the choice $d=2$ that $A_P = \{1,3\}$ and $B_P = \{2,5\}$. Thus, to each d and s, $0 \le d \le s \le n-d$, and P as above we can associate the vector $v_{d,s}(A_P, B_P)$. In our running example of Figure 1 the vector $v_{2,3}(P)$ would hence be $v_{2,3}(\{1,3\},\{2,5\})$, the vector in (2.15). To have a more concise form of notation, we will write $v_{d,s}(P)$ for $v_{d,s}(A_P, B_P)$ from now on.

Lemma 4. *The set of vectors*

$$\{v_{d,s}(P) : 0 \le d \le s \le n-d,\ P \text{ a ballot path from } (0,0) \text{ to } (n, n-2d)\} \tag{2.16}$$

is linearly independent.

The final Lemma tells us how many such vectors $v_{d,s}(P)$ there are.

Lemma 5. *The number of ballot paths from $(0,0)$ to $(n, n-2d)$ is $\frac{n-2d+1}{n+1}\binom{n+1}{d}$. The total number of all vectors in the set (2.16) is 2^n.*

Now, let us for a moment assume that Lemmas 3–5 are already proved. Then, Theorem 2 follows immediately, as it turns out.

Proof of Theorem 2. Consider the set of vectors in (2.16). By Lemma 3 we know that it consists of eigenvectors for the matrix $\zeta_n(u)$. In addition, Lemma 4 tells us that this set of vectors is linearly independent. Furthermore, by Lemma 5 the number of vectors in this set is exactly 2^n, which is the dimension of the space where all these vectors are contained. Therefore, they must form a basis of the space.

Lemma 3 says more precisely that $v_{d,s}(P)$ is an eigenvector for the eigenvalue λ_d. From what we already know, this implies that for fixed d the set

$$\{v_{d,s}(P) : d \le s \le n-d,\ P \text{ a ballot path from } (0,0) \text{ to } (n, n-2d)\}$$

forms a basis for the eigenspace corresponding to λ_d. Therefore, the dimension of the eigenspace corresponding to λ_d equals the number of possible numbers s times the number of possible lattice paths P. This is exactly

$$(n-2d+1)\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d},$$

the number of possible lattice paths P being given by the first statement of Lemma 5. This expression equals exactly the expression (2.13). Thus, Theorem 2 is proved. □

Now we turn to the proofs of the Lemmas.

Proof of Lemma 3. Let d, s and A, B be fixed, satisfying the restrictions in the statement of the Lemma. We have to show that

$$\zeta_n(u) \cdot v_{d,s}(A,B) = \lambda_d v_{d,s}(A,B).$$

Restricting our attention to the I-th component, we see from the definition (2.14) of $v_{d,s}(A,B)$ that we need to establish

$$\sum_{\substack{X \subseteq A\\ Y \subseteq [n]\backslash(A\cup B),\ |Y|=s-d}} Z_{I,X\cup X'\cup Y}\,(-1)^{|X|} = \begin{cases} \lambda_d(-1)^{|U|} & \text{if } I \text{ is of the form } U\cup U'\cup V\\ & \text{for some } U \text{ and } V,\ U\subseteq A,\\ & V\subseteq [n]\backslash(A\cup B),\ |V|=s-d\\ 0 & \text{otherwise.}\end{cases} \tag{2.17}$$

We prove (2.17) by a case by case analysis. The first two cases cover the case "otherwise" in (2.17), the third case treats the first alternative in (2.17).

Case 1. The cardinality of I is different from s. As we observed earlier, the cardinality of any set $X\cup X'\cup Y$ which occurs as index at the left-hand side of (2.17) equals s. The cardinality of I however is different from s. As we observed in the Remark after Theorem 1, this implies that any coefficient $Z_{I,X\cup X'\cup Y}$ on the left-hand side vanishes. Thus, (2.17) is proved in this case.

Case 2. The cardinality of I equals s, but I does not have the form $U\cup U'\cup V$ for any U and V, $U\subseteq A$, $V\subseteq [n]\backslash(A\cup B)$, $|V| = s-d$. Now the sum on the left-hand side of (2.17) contains nonzero contributions. We have to show that they cancel each other. We do this by grouping summands in pairs, the sum of each pair being 0.

Consider a set $X \cup X' \cup Y$ which occurs as index at the left-hand side of (2.17). Let e be minimal such that

either: the e-th largest element of A and the e-th largest element of B are both in I,

or: the e-th largest element of A and the e-th largest element of B are both not in I.

That such an e must exist is guaranteed by our assumptions about I. Now consider X and X'. If the e-th largest element of A is contained in X then the e-th largest element of B is not contained in X', and vice versa. Define a new set $\bar{X}$ by adding to X the e-th largest element of A if it is not already contained in X, respectively by removing it from X if it is contained in X. Then, it is easily checked that

$$Z_{I,X\cup X'\cup Y} = Z_{I,\bar{X}\cup\bar{X}'\cup Y}.$$

On the other hand, we have $(-1)^{|X|} = -(-1)^{|\bar{X}|}$ since the cardinalities of X and $\bar{X}$ differ by ± 1. Both facts combined give

$$Z_{I,X\cup X'\cup Y}\,(-1)^{|X|} + Z_{I,\bar{X}\cup\bar{X}'\cup Y}\,(-1)^{|\bar{X}|} = 0.$$

Hence, we have found two summands on the left-hand side of (2.17) which cancel each other.

Summarizing, this construction finds for any X, Y sets $\bar{X}, Y$ such that the corresponding summands on the left-hand side of (2.17) cancel each other. Moreover, this construction applied to $\bar{X}, Y$ gives back X, Y. Hence, what the construction does is exactly what we claimed, namely it groups the summands into pairs which contribute 0 to the whole sum. Therefore the sum is 0, which establishes (2.17) in this case also.

Case 3. I has the form $U \cup U' \cup V$ for some U and V, $U \subseteq A$, $V \subseteq [n]\backslash(A\cup B)$, $|V| = s - d$. This assumption implies in particular that the cardinality of I is s. From the Remark after the statement of Theorem 1 we know that in our situation $Z_{I,X\cup X'\cup Y}$ depends only on the number of common elements in I and $X \cup X' \cup Y$. Thus, the left-hand side in (2.17) reduces to

$$\sum_{j,k\ge 0} N(j,k)\,(-1)^{|U|+j}\,k!\frac{1}{2^n}\frac{\Gamma\left(\frac{5}{2}-u\right)\,\Gamma(2+n-s-u)\,\Gamma(2+s-u)}{\Gamma\left(\frac{5}{2}+\frac{n}{2}-u\right)\,\Gamma\left(2+\frac{n}{2}-u\right)\Gamma(2+k-u)}, \tag{2.18}$$

where $N(j,k)$ is the number of sets $X \cup X' \cup Y$, for some X and Y, $X \subseteq A$, $Y \subseteq [n]\backslash(A\cup B)$, $|Y| = s-d$, which have $s-k$ elements in common with I, and which have $d-j$ elements in common with $I\cap(A\cup B) = U\cup U'$. Clearly, we used expression (2.2) with $n_{\in\in} = s-k$ and $n_{\notin\notin} = n-s-k$.

To determine $N(j,k)$, note first that there are $\binom{d}{j}$ possible sets $X \cup X'$ which intersect $U \cup U'$ in exactly $d-j$ elements. Next, let us assume that we already made a choice for $X \cup X'$. In order to determine the number of possible sets Y such that $X \cup X' \cup Y$ has $s-k$ elements in common with I, we have to choose $(s-k)-(d-j) = s-d+j-k$ elements from V, for which we have $\binom{s-d}{s-d+j-k}$ possibilities, and we have to choose $s-d-(s-d+j-k) = k-j$ elements from $[n]\backslash(I \cup A \cup B)$ to obtain a total number of s elements, for which we have $\binom{n-s-d}{k-j}$ possibilities. Hence,

$$N(j,k) = \binom{d}{j}\binom{s-d}{k-j}\binom{n-s-d}{k-j}. \tag{2.19}$$

So it remains to evaluate the double sum (2.18), using the expression (2.19) for $N(j,k)$.

We start by writing the sum over j in (2.18) in hypergeometric notation,

$$(-1)^{|U|}\frac{1}{2^n}\frac{\Gamma(\frac{5}{2}-u)\,\Gamma(2+n-s-u)\,\Gamma(2+s-u)}{\Gamma(2-u)\,\Gamma(2+\frac{n}{2}-u)\,\Gamma(\frac{5}{2}+\frac{n}{2}-u)} \\ \times\sum_{k=0}^{\infty}\frac{(d-s)_k\,(d-n+s)_k}{(1)_k\,(2-u)_k}\,{}_3F_2\left[\begin{matrix}-k,-k,-d\\1-d-k+s,1-d-k+n-s\end{matrix};1\right].$$

To the ${}_3F_2$ series we apply a transformation formula of Thomae (see e.g. [25, (3.1.1)]),

$${}_3F_2\left[\begin{matrix}a,b,-m\\d,e\end{matrix};1\right] = \frac{(-b+e)_m}{(e)_m}\,{}_3F_2\left[\begin{matrix}-m,b,-a+d\\d,1+b-e-m\end{matrix};1\right] \tag{2.20}$$

where m is a nonnegative integer. We write the resulting ${}_3F_2$ again as a sum over j, then interchange sums over k and j, and write the (now) inner sum over k in hypergeometric notation. Thus we obtain

$$(-1)^{|U|}\frac{1}{2^n}\frac{\Gamma(\frac{5}{2}-u)\,\Gamma(2+n-s-u)\,\Gamma(2+s-u)}{\Gamma(\frac{5}{2}+\frac{n}{2}-u)\,\Gamma(2+\frac{n}{2}-u)\,\Gamma(2-u)} \\ \times\sum_{j=0}^{\infty}\frac{(-d)_j\,(1-d+s)_j}{(1)_j\,(2-u)_j}\,{}_2F_1\left[\begin{matrix}j-n+s,d-s\\2+j-u\end{matrix};1\right].$$

The ${}_2F_1$ series in this expression is terminating because $d-s$ is a nonpositive integer. Hence, it can be summed by means of Gauß' sum (2.10). Writing the remaining sum over j in hypergeometric notation, the above expression becomes

$$(-1)^{|U|}\frac{1}{2^n}\frac{\Gamma(\frac{5}{2}-u)\,\Gamma(2+n-d-u)\,\Gamma(2+s-u)}{\Gamma(\frac{5}{2}+\frac{n}{2}-u)\,\Gamma(2+\frac{n}{2}-u)\,\Gamma(2+s-d-u)}\,{}_2F_1\left[\begin{matrix}-d,1-d+s\\2-d+s-u\end{matrix};1\right].$$

Again, the ${}_2F_1$ series is terminating and so is summable by means of (2.10). Thus, we get

$$(-1)^{|U|}\frac{1}{2^n}\frac{\Gamma\left(\frac{5}{2}-u\right)\,\Gamma(2+n-d-u)\,\Gamma(1+d-u)}{\Gamma\left(\frac{5}{2}+\frac{n}{2}-u\right)\,\Gamma(2+\frac{n}{2}-u)\,\Gamma(1-u)},$$

which is exactly the expression (2.12) for λ_d times $(-1)^{|U|}$. This proves (2.17) in this case.

The proof of Lemma 3 is now complete. $\square$

Proof of Lemma 4. We know from Lemma 3 that $v_{d,s}(P)$ lies in the eigenspace for the eigenvalue λ_d, with λ_d being given in (2.12). The λ_d's, $d = 0, 1, \dots, \lfloor n/2 \rfloor$, are all distinct, so the corresponding eigenspaces are linearly independent. Therefore it suffices to show that for any *fixed* d the set of vectors

$$\{v_{d,s}(P) : d \le s \le n-d,\ P \text{ a ballot path from } (0,0) \text{ to } (n, n-2d)\}$$

is linearly independent.

On the other hand, a vector $v_{d,s}(A, B)$ lies in the space spanned by the standard unit vectors e_T with $|T| = s$. Clearly, as s varies, these spaces are linearly independent. Therefore, it suffices to show that for any *fixed* d *and* s the set of vectors

$$\{v_{d,s}(P) : P \text{ a ballot path from } (0,0) \text{ to } (n, n-2d)\}$$

is linearly independent.

So, let us fix integers d and s with $0 \le d \le s \le n-d$, and let us suppose that there is some vanishing linear combination

$$\sum_{P \text{ b.p. from } (0,0) \text{ to } (n,n-2d)} c_P\, v_{d,s}(P) = 0. \tag{2.21}$$

We have to establish that $c_P = 0$ for all ballot paths P from $(0,0)$ to $(n, n-2d)$.

We prove this fact by induction on the set of ballot paths from $(0,0)$ to $(n, n-2d)$. In order to make this more precise, we need to impose a certain order on the ballot paths. Given a ballot path P from $(0,0)$ to $(n, n-2d)$, we define its *front portion* F_P to be the portion of P from the beginning up to and including P's d-th up-step. For example, choosing $d = 2$, the front portion of the ballot path in Figure 1 is the subpath from $(0,0)$ to $(3,1)$. Note that F_P can be any ballot path starting in $(0,0)$ with d up-steps and less than d down-steps. We order such front portions lexicographically, in the sense that F_1 is before F_2 if and only if F_1 and F_2 agree up to some point and then F_1 continues with an up-step while F_2 continues with a down-step.

Now, here is what we are going to prove: Fix any possible front portion F. We shall show that $c_P = 0$ for all P with front portion F_P equal to F, *given that it is already known that* $c_{P'} = 0$ *for all* P' *with a front portion* $F_{P'}$ *that is before* F. Clearly, by induction, this would prove $c_P = 0$ for *all* ballot paths P from $(0,0)$ to $(n, n-2d)$.

Let F be a possible front portion, i.e., a ballot path starting in $(0,0)$ with exactly d up-steps and less than d down-steps. As we did earlier, label the steps of F by $1, 2, \dots,$ and denote the set of labels corresponding to the down-steps of F by B_F.

We write b for $|B_F|$, the number of all down-steps of F. Observe that then the total number of steps of F is $d+b$.

Now, let T be a fixed $(d-b)$-element subset of $\{d+b+1, d+b+2, \dots, n\}$. Furthermore, let S be a set of the form $S = B_F \cup S_1 \cup S_2$, where $S_1 \subseteq T$ and $S_2 \subseteq \{d+b+1, d+b+2, \dots, n\} \backslash T$, and such that $|S| = s$.

We consider the coefficient of e_S in the left-hand side of (2.21). To determine this coefficient, we have to determine the coefficient of e_S in $v_{d,s}(P)$, for all P. We may concentrate on those P whose front portion F_P is equal to or later than F, since our induction hypothesis says that $c_P = 0$ for all P with F_P before F. So, let P be a ballot path from $(0,0)$ to $(n, n-2d)$ with front portion equal to or later than F. We claim that the coefficient of e_S in $v_{d,s}(P)$ is zero unless the set B_P of down-steps of P is contained in S.

Let the coefficient of e_S in $v_{d,s}(P)$ be nonzero. To establish the claim, we first prove that the front portion F_P of P has to equal F. Suppose that this is not the case. Then the front portion of P runs in parallel with F for some time, say for the first $(m-1)$ steps, with some $m \le d+b$, and then F continues with an up-step and F_P continues with a down-step (recall that F_P is equal to or later than F). By (2.14) we have

$$v_{d,s}(P) := \sum_{\substack{X \subseteq A_P \\ Y \subseteq [n] \backslash (A_P \cup B_P),\ |Y| = s-d}} (-1)^{|X|} \, e_{X \cup X' \cup Y}. \tag{2.22}$$

We are assuming that the coefficient of e_S in $v_{d,s}(P)$ is nonzero, therefore S must be of the form $S = X \cup X' \cup Y$, with X, Y as described in (2.22). We are considering the case that the m-th step of F_P is a down-step, whence $m \in B_P$, while the m-th step of F is an up-step, whence $m \notin B_F$. By definition of S, we have $S \cap \{1, 2 \dots, d+b\} = B_F$, whence $m \notin S$.

Summarizing so far, we have $m \in B_P$, $m \notin S$, for some $m \le d+b$, and $S = X \cup X' \cup Y$, for some X, Y as described in (2.22). In particular we have $m \notin X'$. Now recall that X' is the "complement of X in B_P". This says in particular that, if m is the i-th largest element in B_P, then the i-th largest element of A_P, a say, is an element of X, and so of S. By construction of A_P and B_P, a is smaller than m, so in particular $a < d+b$. As we already observed, there holds $S \cap \{1, 2, \dots, d+b\} = B_F$, so we have $a \in B_F$, i.e., the a-th step of F is a down-step. On the other hand, we assumed that P and F run in parallel for the first $(m-1)$ steps. Since $a \in A_P$, the set of up-steps of P, the a-th step of P is an up-step. We have $a \le m-1$, therefore the a-th step of F must be an up-step also. This is absurd. Therefore, given that the coefficient of e_S in $v_{d,s}(P)$ is nonzero, the front portion F_P of P has to equal F.

Now, let P be a ballot path from $(0,0)$ to $(n, n-2d)$ with front portion equal to F, and suppose that S has the form $S = X \cup X' \cup Y$, for some X, Y as described in (2.22). By definition of the front portion, the set A_P of up-steps of P has the property $A_P \cap \{1,2,\dots,d+b\} = \{1,2,\dots,d+b\}\backslash B_F$. Since $|B_F| = b$, these are the labels of exactly d up-steps. Since the cardinality of A_P is exactly d by definition, we must have $A_P = \{1,2,\dots,d+b\}\backslash B_F$. Because of $S \cap \{1,2,\dots,d+b\} = B_F$, which we already used a number of times, A_P and S are disjoint, which in particular implies that A_P and X are disjoint. However, X is a subset of A_P by definition, so X must be empty. This in turn implies that $X' = B_P$. This says nothing else but that the set B_P of down-steps of P equals X' and so is contained in S. This establishes our claim.

In fact, we proved more. We saw that S has the form $S = X \cup X' \cup Y$, with $X = \emptyset$. This implies that the coefficient of e_S in $v_{d,s}(P)$, as given by (2.22), is actually $+1$. Comparison of coefficients of e_S in (2.21) then gives

$$\sum_{\substack{P \text{ b.p. from } (0,0) \text{ to } (n,n-2d)\\ F_P=F,\ B_P\subseteq S}} c_P = 0, \tag{2.23}$$

for any $S = B_F \cup S_1 \cup S_2$, where $S_1 \subseteq T$ and $S_2 \subseteq \{d+b+1, d+b+2, \dots, n\}\backslash T$, and such that $|S| = s$.

Now, we sum both sides of (2.23) over all such sets S, keeping the cardinality of S_1 and S_2 fixed, say $|S_1| = d-b-j$, enforcing $|S_2| = s-d+j$, for a fixed j, $0 \le j \le d-b$. For a fixed ballot path P from $(0,0)$ to $(n, n-2d)$, with front portion F, with $d-b-k$ down-steps in T, and hence with k down-steps in $\{d+b+1, d+b+2, \dots, n\}\backslash T$, there are $\binom{k}{k-j}$ such sets $S_1 \subseteq T$ containing all the $d-b-k$ down-steps of P in T, and there are $\binom{n-(d+b)-(d-b)-k}{s-d+j-k}$ such sets $S_2 \subseteq \{d+b+1, d+b+2, \dots, n\}\backslash T$ containing all the k down-steps of P in $\{d+b+1, d+b+2, \dots, n\}\backslash T$. Therefore, summing up (2.23) gives

$$\sum_{k\ge 0} \binom{k}{j}\binom{n-2d-k}{n-d-s-j}\Bigg(\sum_{\substack{P \text{ b.p. from } (0,0) \text{ to } (n,n-2d)\\ F_P=F,\ |B_P\cap T|=d-b-k\\ |B_P\cap(\{d+b+1,d+b+2,\dots,n\}\backslash T)|=k}} c_P\Bigg) = 0, \quad j = 0,1,\dots,d-b. \tag{2.24}$$

Denoting the inner sum in (2.24) by $C(k)$, we see that (2.24) represents a non-degenerate triangular system of linear equations for $C(0), C(1), \dots, C(d-b)$. Therefore, all the quantities $C(0), C(1), \dots, C(d-b)$ have to equal 0. In particular, we have $C(0) = 0$. Now, $C(0)$ consists of just a single term c_P, with P being the ballot path from $(0,0)$ to $(n, n-2d)$, with front portion F, and the labels of the $d-b$ down-steps besides those of F being exactly the elements of T. Therefore, we have $c_P = 0$ for this ballot path. The set T was an arbitrary $(d-b)$-subset of $\{d+b+1, d+b+2, \dots, n\}$. Thus, we have proved $c_P = 0$ for any ballot path P from $(0,0)$ to $(n, n-2d)$ with front portion F. This completes our induction proof. $\square$

Proof of Lemma 5. That the number of ballot paths from $(0,0)$ to $(n,n-2d)$ equals $\frac{n-2d+1}{n+1}\binom{n+1}{d}$ is a classical combinatorial result (see e.g. [38, Theorem 1 with $t=1$]). From this it follows that the total number of vectors in the set (2.16) is

$$\sum_{d=0}^{\lfloor n/2\rfloor}(n-2d+1)\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}. \tag{2.25}$$

To evaluate this sum, note that the summand is invariant under the substitution $d\to n-2d+1$. Therefore, extending the range of summation in (2.25) to $d=0,1,\dots,n+1$ and dividing the result by 2 gives the same value. So, the cardinality of the set (2.16) is also given by

$$\frac{1}{2}\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}.$$

Using the simple identity

$$\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}=(n+1)\binom{n+1}{d}-4n\binom{n}{d-1}+4n\binom{n-1}{d-2},$$

the last sum can be decomposed into

$$\frac{n+1}{2}\sum_{d=0}^{n+1}\binom{n+1}{d}-2n\sum_{d=1}^{n+1}\binom{n}{d-1}+2n\sum_{d=2}^{n+1}\binom{n-1}{d-2}.$$

Each of these sums can be evaluated by the binomial theorem, and thus the expression reduces to 2^n. This completes the proof of the Lemma. $\square$

In fact, Theorem 2 can be generalized to a wider class of matrices.

Theorem 6. *Let $\tilde\zeta_n(u)=(\tilde Z_{IJ})_{I,J\in[n]}$ be the $2^n\times 2^n$ matrix defined by*

$$\tilde Z_{IJ}:=\delta_{n_{\notin\in},n_{\in\notin}}\frac{\left(\frac{n-n_{\in\in}-n_{\notin\notin}}{2}\right)!}{\Gamma\left(2+\frac{n-n_{\in\in}-n_{\notin\notin}}{2}-u\right)}\cdot f(n_{\in\in}-n_{\notin\notin}),$$

where $n_{\in\in}$, etc., have the same meaning as earlier, and where $f(x)$ is a function of x which is symmetric, i.e., $f(x)=f(-x)$. Then, the eigenvalues of $\tilde\zeta_n(u)$ are

$$\lambda_{d,s}=f(n-2s)\frac{\Gamma(2+n-d-u)\,\Gamma(1+d-u)}{\Gamma(2+n-s-u)\,\Gamma(2+s-u)\,\Gamma(1-u)},\quad 0\le d\le s\le n-d, \tag{2.26}$$

with respective multiplicities

$$\frac{n-2d+1}{n+1}\binom{n+1}{d}, \tag{2.27}$$

independent of s.

PROOF. The above proof of Theorem 2 has to be adjusted only insignificantly to yield a proof of Theorem 6. In particular, the vector $v_{d,s}(A,B)$ as defined in (2.14) is an eigenvector for $\lambda_{d,s}$, for any two disjoint d-element subsets A and B of $[n]$, and the set (2.16) is a basis of eigenvectors for $\tilde{\zeta}_n(u)$. $\square$

2.3. The relative entropies of $\overset{n}{\otimes}\rho$ with respect to the Bayesian density matrices $\zeta_n(u)$.

We now apply the preceding results to compute the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$. Utilizing the definition (1.5) of relative entropy and employing the property [39, 58] that $S(\overset{n}{\otimes}\rho) = nS(\rho)$, it is given by

$$-n\,S(\rho) - \operatorname{Tr}\left(\overset{n}{\otimes}\rho \cdot \log \zeta_n(u)\right). \tag{2.28}$$

The term $S(\rho)$ has been given in (1.7). Concerning the second term in (2.28), we have the following theorem.

Theorem 7. *Let $\zeta_n(u) = (Z_{IJ})_{I,J\in[n]}$ be the matrix with entries Z_{IJ} given in* (2.2). *Then, we have*

$$\begin{aligned}&\operatorname{Tr}\left(\overset{n}{\otimes}\rho \cdot \log \zeta_n(u)\right)\\ &= \sum_{d=0}^{\lfloor n/2\rfloor} \frac{n-2d+1}{n+1}\binom{n+1}{d}\frac{1}{2^{n+1}r}\left((1+r)^{n+1-d}(1-r)^d - (1+r)^d(1-r)^{n+1-d}\right)\log\lambda_d,\end{aligned} \tag{2.29}$$

with λ_d as given in (2.12).

Before we move on to the proof, we note that Theorem 7 gives us the following expression for the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$

Corollary 8. *The relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$ equals*

$$\begin{aligned}&\frac{n}{2}(1-r)\log((1-r)/2) + \frac{n}{2}(1+r)\log((1+r)/2)\\ &\qquad - \sum_{d=0}^{\lfloor n/2\rfloor}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}\\ &\qquad\quad \cdot \frac{1}{2^{n+1}r}\left((1+r)^{n-d+1}(1-r)^d - (1+r)^d(1-r)^{n-d+1}\right)\log\lambda_d,\end{aligned} \tag{2.30}$$

with λ_d as given in (2.12).

PROOF OF THEOREM 7. One way of determining the trace of a linear operator L is to choose a basis of the vector space, $\{v_I : I \in [n]\}$ say, write the action of L on the basis elements in the form

$$Lv_I = c_I v_I + \text{linear combination of } v_J\text{'s},\ J \neq I,$$

and then form the sum $\sum_I c_I$ of the "diagonal" coefficients, which gives exactly the trace of L.

Clearly, we choose as a basis our set (2.16) of eigenvectors for $\zeta_n(u)$. To determine the action of $\overset{n}{\otimes}\rho \cdot \log \zeta_n(u)$ we need only to find the action of $\overset{n}{\otimes}\rho$ on the vectors in the set (2.16). We claim that this action can be described as

$$\begin{aligned}\big(\overset{n}{\otimes}\,\rho\big) \cdot v_{d,s}(P) &\\ = \frac{1}{2^n}\Bigg(\sum_{k\ge j\ge 0}(-1)^j\binom{d}{j}\binom{s-d}{k-j}\binom{n-s-d}{k-j}&(1+z)^{s-k}(x^2+y^2)^k(1-z)^{n-s-k}\Bigg)\\ \cdot\, v_{d,s}(P)\ +\ \text{linear combination of eigenvectors}&\\ v_{d',s'}(P') \text{ with } s' \neq s,&\end{aligned} \tag{2.31}$$

for any basis vector $v_{d,s}(P)$ in (2.16).

To see this, consider the I-th component of $\big(\overset{n}{\otimes}\,\rho\big)\cdot v_{d,s}(P)$, i.e., the coefficient of e_I in $\big(\overset{n}{\otimes}\,\rho\big)\cdot v_{d,s}(P)$, $I \in [n]$. By the definition (2.14) of $v_{d,s}(P)$ it equals

$$\sum_{\substack{X\subseteq A_P\\ Y\subseteq[n]\setminus(A_P\cup B_P),\ |Y|=s-d}} R_{I,X\cup X'\cup Y}\,(-1)^{|X|}, \tag{2.32}$$

where R_{IJ} denotes the (I,J)-entry of $\overset{n}{\otimes}\rho$. (Recall that R_{IJ} is given explicitly in (2.1).) Now, it should be observed that we did a similar calculation already, namely in the proof of Lemma 3. In fact, the expression (2.32) is almost identical with the left-hand side of (2.17). The essential difference is that Z_{IJ} is replaced by R_{IJ} for all J (the nonessential difference is that A, B are replaced by A_P, B_P, respectively). Therefore, we can partially rely upon what was done in the proof of Lemma 3.

We distinguish between the same cases as in the proof of Lemma 3.

Case 1. The cardinality of I is different from s. We do not have to worry about this case, since e_I then lies in the span of vectors $v_{d',s'}(P')$ with $s' \neq s$, which is taken care of in (2.31).

Case 2. The cardinality of I equals s, but I does not have the form $U \cup U' \cup V$ for any U and V, $U \subseteq A_P$, $V \subseteq [n]\setminus(A_P \cup B_P)$, $|V| = s-d$. Essentially the same arguments as those in Case 2 in the proof of Lemma 3 show that the term (2.32) vanishes for this choice of I. Of course, one has to use the explicit expression (2.1) for R_{IJ}.

Case 3. I has the form $U\cup U'\cup V$ for some U and V, $U \subseteq A_P$, $V \subseteq [n]\setminus(A_P\cup B_P)$, $|V| = s-d$. In Case 3 in the proof of Lemma 3 we observed that there are $N(j,k)$

sets $X \cup X' \cup Y$, for some X and Y, $X \subseteq A_P$, $Y \subseteq [n]\backslash(A_P \cup B_P)$, $|Y| = s-d$, which have $s-k$ elements in common with I, and which have $d-j$ elements in common with $I \cap (A_P \cup B_P) = U \cup U'$, where $N(j,k)$ is given by (2.19). Then, using the explicit expression (2.1) for R_{IJ}, it is straightforward to see that the expression (2.32) equals

$$\frac{1}{2^n}\sum_{k\ge j\ge 0}(-1)^{|U|+j}\binom{d}{j}\binom{s-d}{k-j}\binom{n-s-d}{k-j}(1+z)^{s-k}(x^2+y^2)^k(1-z)^{n-s-k}$$

in this case. This establishes (2.31).

Now we are in the position to write down an expression for the trace of $\overset{n}{\otimes}\rho\cdot\log\zeta_n(u)$. By Theorem 2 and by (2.31) we have

$$\begin{aligned}\Big(\overset{n}{\otimes}\rho\cdot\log\zeta_n(u)\Big)&\cdot v_{d,s}(P)\\ =\frac{1}{2^n}&\Bigg(\sum_{k\ge j\ge 0}(-1)^j\binom{d}{j}\binom{s-d}{k-j}\binom{n-s-d}{k-j}(1+z)^{s-k}(x^2+y^2)^k(1-z)^{n-s-k}\Bigg)\\ &\cdot\log\lambda_d\cdot v_{d,s}(P)+\text{linear combination of eigenvectors}\\ &\qquad v_{d',s'}(P')\text{ with }s'\ne s.\end{aligned}\tag{2.33}$$

From what was said at the beginning of this proof, in order to obtain the trace of $\overset{n}{\otimes}\rho\cdot\log\zeta_n(u)$, we have to form the sum of all the "diagonal" coefficients in (2.33). Using the first statement of Lemma 5 and replacing x^2+y^2 by r^2-z^2, we see that it is

$$\begin{aligned}\sum_{d=0}^{\lfloor n/2\rfloor}\log\lambda_d\,\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}\frac{1}{2^n}\sum_{s=d}^{n-d}\sum_{k\ge j\ge 0}(-1)^j\binom{d}{j}\binom{s-d}{k-j}\binom{n-s-d}{k-j}\\ \cdot(1+z)^{s-k}(r^2-z^2)^k(1-z)^{n-s-k}.\end{aligned}\tag{2.34}$$

In order to see that this expression equals (2.29), we have to prove

$$\begin{aligned}\sum_{s=d}^{n-d}\sum_{j=0}^{d}\sum_{k=j}^{s}(-1)^j\binom{d}{j}\binom{s-d}{k-j}\binom{n-s-d}{k-j}(1+z)^{s-k}(r^2-z^2)^k(1-z)^{n-s-k}\\ =\frac{1}{2r}\big((1+r)^{n+1-d}(1-r)^d-(1+r)^d(1-r)^{n+1-d}\big).\end{aligned}\tag{2.35}$$

We start with the left-hand side of (2.35) and write the inner sum in hypergeometric notation, thus obtaining

$$\sum_{s=d}^{n-d}\sum_{j=0}^{d}(1-z)^{n-s-j}(1+z)^{s-j}\big(r^2-z^2\big)^j\frac{(-d)_j}{(1)_j}{}_2F_1\left[\begin{matrix}d-n+s,d-s\\1\end{matrix};\frac{r^2-z^2}{1-z^2}\right].$$

To the ${}_2F_1$ series we apply the transformation formula ([48, (1.8.10), terminating form]

$${}_2F_1\left[\begin{matrix}a,-m\\ c\end{matrix};z\right]=\frac{(c-a)_m}{(c)_m}\,{}_2F_1\left[\begin{matrix}-m,a\\ 1+a-c-m\end{matrix};1-z\right],$$

where m is a nonnegative integer. We write the resulting ${}_2F_1$ series again as a sum over k. In the resulting expression we exchange sums so that the sum over j becomes the innermost sum. Thus, we obtain

$$\sum_{s=d}^{n-d}\sum_{k=0}^{s-d}\left(1-r^2\right)^k(1-z)^{n-s-k}(1+z)^{s-k}\\ \cdot\frac{(d-s)_k\,(n-d-s+1)_{s-d}\,(d-n+s)_k}{(1)_k\,(1)_{s-d}\,(2d-n)_k}\sum_{j=0}^{d}\binom{d}{j}\left(\frac{z^2-r^2}{1-z^2}\right)^j.$$

Clearly, the innermost sum can be evaluated by the binomial theorem. Then, we interchange sums over s and k. The expression that results is

$$\sum_{k=0}^{\lfloor n/2\rfloor-d}\left(1-r^2\right)^{d+k}(1-z)^{n-2d-2k}\frac{(2d+k-n)_k}{(1)_k}\\ \cdot\sum_{s=0}^{n-2d-2k}\binom{n-2d-2k}{s}\left(\frac{1+z}{1-z}\right)^s.$$

Again, we can apply the binomial theorem. Thus, we reduce our expression on the left-hand side of (2.35) to

$$2^{n-2d}\left(1-r^2\right)^d\sum_{k=0}^{\lfloor n/2\rfloor-d}\frac{\left(d-\frac{n}{2}\right)_k\,\left(d-\frac{n}{2}+\frac{1}{2}\right)_k}{(2d-n)_k\,k!}(1-r^2)^k.$$

Now, we replace $(1-r^2)^k$ by its binomial expansion $\sum_{l=0}^{k}(-1)^l\binom{k}{l}r^{2l}$, interchange sums over k and l, and write the (now) inner sum over k in hypergeometric notation. This gives

$$2^{n-2d}\left(1-r^2\right)^d\Bigg(\sum_{l=0}^{\lfloor n/2\rfloor-d}(-1)^l r^{2l}\frac{(d-\frac{n}{2})_l\,(\frac{1}{2}+d-\frac{n}{2})_l}{(1)_l\,(2d-n)_l}\\ \cdot{}_2F_1\left[\begin{matrix}d+l-\frac{n}{2},\frac{1}{2}+d+l-\frac{n}{2}\\ 2d+l-n\end{matrix};1\right]\Bigg).$$

Finally, this ${}_2F_1$ series can be summed by means of Gauß' summation (2.10). Simplifying, we have

$$\left(1-r^2\right)^d \sum_{l=0}^{\lfloor n/2\rfloor-d} \binom{n-2d+1}{2l+1} r^{2l},$$

which is easily seen to equal the right-hand side in (2.35). This completes the proof of the Theorem. $\square$

2.4. **Asymptotics of the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$.** In the preceding subsection, we obtained in Corollary 8 the general formula (2.30) for the relative entropy of $\overset{n}{\otimes}\rho$ with respect to the Bayesian density matrix $\zeta_n(u)$. We, now, proceed to find its asymptotics for $n\to\infty$. We prove the following theorem.

Theorem 9. *The asymptotics of the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$ for a fixed r with $0\le r<1$ is given by*

$$\frac{3}{2}\log n-\frac{1}{2}-\frac{3}{2}\log 2-(1-u)\log(1-r^2)+\frac{1}{2r}\log\left(\frac{1-r}{1+r}\right)\\ +\log\Gamma(1-u)-\log\Gamma(5/2-u)+O\left(\frac{1}{n}\right). \quad (2.36)$$

In the case $r=0$, this means that the asymptotics is given by the expression (2.36) *in the limit $r\downarrow 0$, i.e., by*

$$\frac{3}{2}\log n-\frac{3}{2}-\frac{3}{2}\log 2+\log\Gamma(1-u)-\log\Gamma(5/2-u)+O\left(\frac{1}{n}\right). \quad (2.37)$$

For any fixed $\varepsilon>0$, the $O(.)$ term in (2.36) *is uniform in u and r as long as $0\le r\le 1-\varepsilon$.*

For $r=1$ the asymptotics is given by

$$(2-u)\log n+(2u-3)\log 2+\frac{1}{2}\log\pi-\log\Gamma(5/2-u)+O\left(\frac{1}{n}\right). \quad (2.38)$$

Also here, the $O(.)$ term is uniform in u.

Remark. It is instructive to observe that, although a comparison of (2.36) and (2.38) seems to suggest that the asymptotics of the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$ behaves completely differently for $0\le r<1$ and $r=1$, the two cases are really quite compatible. In fact, letting r tend to 1 in (2.36) shows that (ignoring the error term) the asymptotic expression approaches $+\infty$ for $u<1/2$, $-\infty$ for $u>1/2$, and it approaches $\frac{3}{2}\log n-\frac{1}{2}-\frac{5}{2}\log 2+\frac{1}{2}\log\pi$ for $u=1/2$. This indicates that, for $r=1$, the order of magnitude of the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$ should be larger than $\frac{3}{2}\log n$ if $u<1/2$, smaller than $\frac{3}{2}\log n$ if $u>1/2$, and exactly $\frac{3}{2}\log n$ if

$u = 1/2$. How much larger or smaller is precisely what formula (2.38) tells us: the order of magnitude is $(2-u)\log n$, and in the case $u = 1/2$ the asymptotics is, in fact, $\frac{3}{2}\log n - 2\log 2 + \frac{1}{2}\log \pi$.

The proof of Theorem 9 relies on several auxiliary summations and estimations. These are stated and proved separately in Lemma 10 and 11.

PROOF OF THEOREM 9. We start with the case $0 < r < 1$. We concentrate first on the sum in (2.30). Because of $\lambda_{n+1-d} = \lambda_d$, we have

$$\frac{1}{2^{n+1}r}\sum_{d=0}^{\lfloor n/2\rfloor}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}\\ \cdot\big((1+r)^{n-d+1}(1-r)^d-(1+r)^d(1-r)^{n-d+1}\big)\log\lambda_d\\ =\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n-d+1}(1-r)^d\log\lambda_d.$$

We expand the logarithm according to the addition rule to obtain

$$\frac{1}{2^{n+1}r}\sum_{d=0}^{\lfloor n/2\rfloor}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}\\ \cdot\big((1+r)^{n-d+1}(1-r)^d-(1+r)^d(1-r)^{n-d+1}\big)\log\lambda_d\\ =\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n-d+1}(1-r)^d\\ \cdot\log\left(\frac{1}{2^n}\frac{\Gamma(5/2-u)}{\Gamma(5/2+n/2-u)\,\Gamma(2+n/2-u)\,\Gamma(1-u)}\right)\\ +\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n-d+1}(1-r)^d\log\Gamma(1+d-u)\\ -\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1-r)^{n-d+1}(1+r)^d\log\Gamma(1+d-u). \tag{2.39}$$

The first sum on the right-hand side of (2.39) can be evaluated by means of (2.41). Besides, by Stirling's formula we have

$$\log\Gamma(z)=\left(z-\frac{1}{2}\right)\log(z)-z+\frac{1}{2}\log 2+\frac{1}{2}\log\pi+O\left(\frac{1}{z}\right).$$

Thus, we get

$$
\begin{aligned}
&\frac{1}{2^{n+1}r}\sum_{d=0}^{\lfloor n/2\rfloor}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d} \\
&\qquad\qquad\cdot\big((1+r)^{n-d+1}(1-r)^{d}-(1+r)^{d}(1-r)^{n-d+1}\big)\log\lambda_d \\
&= -n\log 2-\log\Gamma(5/2+n/2-u)-\log\Gamma(2+n/2-u)+\log\Gamma(5/2-u) \\
&\quad -\log\Gamma(1-u)+\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n-d+1}(1-r)^{d} \\
&\cdot\left((1/2-u+d)\log(1+d-u)-(1-u+d)+\frac{1}{2}\log 2+\frac{1}{2}\log\pi+O\left(\frac{1}{d+1}\right)\right) \\
&\quad -\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1-r)^{n-d+1}(1+r)^{d} \\
&\cdot\left((1/2-u+d)\log(1+d-u)-(1-u+d)+\frac{1}{2}\log 2+\frac{1}{2}\log\pi+O\left(\frac{1}{d+1}\right)\right).
\end{aligned}
\tag{2.40}
$$

Now, the sums are split into several sums by additivity. Those which arise from the first sum in (2.40) can be evaluated using (2.41), (2.42), (2.43), or approximated using (2.48). Those which arise from the second sum can be evaluated by the same identities and approximations, only with r replaced by its negative. Thus, we obtain

$$
\begin{aligned}
&\frac{1}{2^{n+1}r}\sum_{d=0}^{\lfloor n/2\rfloor}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d} \\
&\qquad\qquad\cdot\big((1+r)^{n-d+1}(1-r)^{d}-(1+r)^{d}(1-r)^{n-d+1}\big)\log\lambda_d \\
&= \frac{n}{2}(1-r)\log((1-r)/2)+\frac{n}{2}(1+r)\log((1+r)/2)-\frac{3}{2}\log n+\frac{3}{2}\log 2+\frac{1}{2} \\
&\qquad +(1-u)\log(1-r)+(1-u)\log(1+r)+\frac{1}{2r}\log\left(\frac{1+r}{1-r}\right) \\
&\qquad\qquad +\log\Gamma(5/2-u)-\log\Gamma(1-u)+O\left(\frac{1}{n}\right).
\end{aligned}
$$

Finally, use of this in (2.30) gives the claimed asymptotics (2.36) for the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$.

A closer analysis of the error terms shows that they can, in fact, be bounded uniformly in u and r, $0 < r \le 1-\varepsilon$, for any fixed positive ε.

Now we turn to the two exceptional cases $r=0$ and $r=1$.

In the case $r=1$, by (2.30) the relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$ equals

$$\frac{n}{2}(1-r)\log((1-r)/2)+\frac{n}{2}(1+r)\log((1+r)/2)-\log\lambda_0,$$

λ_0 being given by (2.12). A straightforward application of Stirling's formula then leads to (2.38).

In the case $r=0$, the relative entropy (2.30) of $\overset{n}{\otimes}\rho$ with respect to $\zeta_n(u)$ reduces to

$$\frac{n}{2}(1-r)\log((1-r)/2)+\frac{n}{2}(1+r)\log((1+r)/2)\\ -\frac{1}{2^n}\sum_{d=0}^{\lfloor n/2\rfloor}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\log\lambda_d.$$

The asymptotics of that expression can be determined in a similar way to what was done for $0<r<1$. For the sake of brevity, we omit the derivation. The result is (2.37). Actually, it is possible to rearrange the computations that we did for $0<r<1$, so that in the limit $r\downarrow 0$ they give a proof of (2.37). This last observation justifies the claim that the error term in (2.36) is uniform in u and r, $0\le r\le 1-\varepsilon$ (i.e., including $r=0$), for any fixed positive ε.

This completes the proof of the Theorem. □

Now, we list the summations which were used in the proof of the Theorem.

Lemma 10. *We have the following summations:*

$$\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^d=1. \tag{2.41}$$

$$\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^d d=\frac{(1-r)(nr-1)}{2r}. \tag{2.42}$$

$$\frac{1}{2^{n+1}r}\sum_{d=-1}^{n+1}\frac{(n-2d+1)}{(n+1)(n+2)}\binom{n+2}{d+1}(1+r)^{n+1-d}(1-r)^d=\frac{2(1+2r+nr)}{(n+1)(n+2)r(1-r)}. \tag{2.43}$$

$$\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^d(1/2-u+d)\\ =\frac{-1+2r+nr-nr^2-2ru}{2r}. \tag{2.44}$$

$$\begin{aligned}\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^{d}&\\ \cdot(1/2-u+d)\left(1+d-u-\frac{n(1-r)}{2}\right)&\\ =\frac{-5-n+7r+5nr-3nr^2-nr^3+4u-10ru-2nru+2nr^2u+4ru^2}{4r}.&\end{aligned} \tag{2.45}$$

$$\begin{aligned}&\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^{d}\\ &\qquad\cdot(1/2-u+d)\left(1+d-u-\frac{n(1-r)}{2}\right)^2\\ &=\frac{1}{8r}(-22-9n+26r+24nr+n^2r-5nr^2-n^2r^2-8nr^3-n^2r^3-2nr^4+n^2r^4+32u\\ &+4nu-48ru-22nru+12nr^2u+6nr^3u-12u^2+32ru^2+4nru^2-4nr^2u^2-8ru^3).\end{aligned} \tag{2.46}$$

$$\begin{aligned}&\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^{d}\\ &\qquad\cdot(1/2-u+d)\left(1+d-u-\frac{n(1-r)}{2}\right)^3\\ &=\frac{1}{16r}(-92-61n-3n^2+100r+105nr+15n^2r+19nr^2-4n^2r^2-35nr^3-20n^2r^3-22nr^4\\ &+7n^2r^4-6nr^5+5n^2r^5+188u+60nu-228ru-162nru-6n^2ru+20nr^2u\\ &+6n^2r^2u+66nr^3u+6n^2r^3u+16nr^4u-6n^2r^4u-132u^2-12nu^2+204ru^2\\ &+72nru^2-36nr^2u^2-24nr^3u^2+32u^3-88ru^3-8nru^3+8nr^2u^3+16ru^4).\end{aligned} \tag{2.47}$$

Proof. In all the cases, the sums can be split into several simpler sums, each of which can itself be summed using the binomial theorem. □

Lemma 11. *For fixed r with $0<r<1$ we have the following asymptotic expansion:*

$$\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^d(1/2-u+d)\log(1+d-u)\\
=\left(\frac{n}{2}(1-r)+1-u-\frac{1}{2r}\right)\big(\log n+\log(1-r)-\log 2\big)\\
+\frac{7}{4}-u+\frac{r}{4}-\frac{1}{2r}+O\left(\frac{1}{n}\right). \qquad (2.48)$$

Proof. We start with the expansion

$$\log(1+d-u)=\log\left(\frac{n(1-r)}{2}\right)+\log\left(1+\frac{2}{n(1-r)}\left(1+d-u-\frac{n(1-r)}{2}\right)\right)\\
=\log n+\log(1-r)-\log 2+\frac{2}{n(1-r)}\left(1+d-u-\frac{n(1-r)}{2}\right)\\
-\frac{2}{n^2(1-r)^2}\left(1+d-u-\frac{n(1-r)}{2}\right)^2\\
+O\left(\frac{1}{n^3(1-r)^3}\left(1+d-u-\frac{n(1-r)}{2}\right)^3\right).$$

(It is at this point that we must have $r<1$.) If we use this expansion in the left-hand side of (2.48) and subsequently use (2.44)–(2.47) to evaluate the resulting sums, we obtain

$$\frac{1}{2^{n+1}r}\sum_{d=0}^{n+1}\frac{(n-2d+1)}{(n+1)}\binom{n+1}{d}(1+r)^{n+1-d}(1-r)^d(1/2-u+d)\log(1+d-u)\\
=\left(\frac{n}{2}(1-r)+1-u-\frac{1}{2r}\right)\big(\log n+\log(1-r)-\log 2\big)\\
+\left(2-u+\frac{r}{2}-\frac{1}{2r}\right)-\frac{1+r}{4}+O\left(\frac{1}{n}\right). \qquad (2.49)$$

Simplifying easily, we obtain (2.48). $\square$

2.5. **Asymptotics of the von Neumann entropies of the Bayesian density matrices $\zeta_n(u)$.** The main result of this section describes the asymptotics of the von Neumann entropy (1.1) of $\zeta_n(u)$. In view of the explicit description of the eigenvalues of $\zeta_n(u)$ and their multiplicities in Theorem 2, this entropy equals

$$-\sum_{d=0}^{\lfloor n/2\rfloor}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d\log\lambda_d,$$

with λ_d being given by (2.12).

Theorem 12. *We have the following asymptotic expansion:*

$$\begin{aligned}-\sum_{d=0}^{\lfloor n/2\rfloor}&\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d\log\lambda_d=\\ &n\left(\frac{-7+5u}{2\,(2-u)\,(1-u)}+\psi(5-2u)-\psi(1-u)\right)+\frac{3}{2}\log n+\left(-\frac{7}{2}+2u\right)\log 2\\ &\qquad-\frac{14-20u+7u^2}{2\,(2-u)\,(1-u)}+\log\big(\Gamma(1-u)\big)-\log\big(\Gamma(5/2-u)\big)\\ &\qquad\qquad+(2-2u)(\psi(5-2u)-\psi(1-u))+O\left(\frac{1}{n^{1-u}}\right),\end{aligned}\tag{2.50}$$

where $\psi(x)$ is the digamma function,

$$\psi(x)=\frac{\frac{d}{dx}\Gamma(x)}{\Gamma(x)}.$$

The proof of the Theorem depends on a few summations, which we now list.

Lemma 13. *We have the following summations:*

$$\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d=2.\tag{2.51}$$

$$\sum_{d=1}^{n+1}(n-2d+1)^2\binom{n}{d-1}\lambda_d=n+1.\tag{2.52}$$

$$\sum_{d=-1}^{n+1}\frac{(n-2d+1)^2}{(n+1)(n+2)}\binom{n+2}{d+1}\lambda_d=\frac{2(n+3)(2u-3)}{(n+1)(n+2)u}.\tag{2.53}$$

$$\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}$$
$$\cdot\frac{1}{2^n}\frac{\Gamma(5/2-u)\,\Gamma(2+n-d-u)\,\Gamma(1+\alpha+d-u)}{\Gamma(5/2+n/2-u)\,\Gamma(2+n/2-u)\,\Gamma(1-u)}(d-u+1/2)$$
$$=(48+64\alpha+25\alpha^2+3\alpha^3+40n+66\alpha n+37\alpha^2n+5\alpha^3n+8n^2+14\alpha n^2+8\alpha^2n^2$$
$$+2\alpha^3n^2-152u-138\alpha u-34\alpha^2u-2\alpha^3u-92nu-92\alpha nu-32\alpha^2nu-2\alpha^3nu$$
$$-12n^2u-10\alpha n^2u-2\alpha^2n^2u+176u^2+100\alpha u^2+12\alpha^2u^2+68nu^2+32\alpha nu^2$$
$$+4\alpha^2nu^2+4n^2u^2-88u^3-24\alpha u^3-16nu^3+16u^4)$$
$$\times\frac{\Gamma(5-2u)\,\Gamma(3+\alpha+n-2u)\,\Gamma(1+\alpha-u)}{4\,\Gamma(5+\alpha-2u)\,\Gamma(4+n-2u)\,\Gamma(3-u)}. \quad (2.54)$$

$$\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d(d-u+1/2)\psi(1+d-u)$$
$$=\frac{32+33n+7n^2-69u-46nu-5n^2u+50u^2+16nu^2-12u^3}{2\,(2-u)\,(1-u)\,(3+n-2u)}$$
$$+(n+2-2u)(\psi(1-u)+\psi(n+3-2u)-\psi(5-2u)). \quad (2.55)$$

Proof. Identities (2.51), (2.52), (2.53), (2.54) are proved by splitting the sums appropriately so that each part can be summed by means of Gauß' ${}_2F_1$ summation. Identity (2.55) follows from (2.54) by differentiating with respect to α and then setting $\alpha=0$. □

From (2.55) we can deduce the following important estimation. The result and its proof were kindly reported to us by Peter Grabner.

Lemma 14. *We have the asymptotic expansion:*

$$\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d(d-u+1/2)\log(1+d-u)$$
$$=n\left(\log n+\frac{7-5u}{2\,(2-u)\,(1-u)}-\psi(5-2u)+\psi(1-u)\right)+(2-2u)\log n$$
$$+\frac{26-46u+25u^2-4u^3}{2\,(2-u)\,(1-u)}+(-2+2u)\,\psi(5-2u)+(2-2u)\,\psi(1-u)+O\left(\frac{1}{n^{1-u}}\right).$$
$$(2.56)$$

Proof. We use the asymptotic expansion

$$\psi(z) = \log(z) - \frac{1}{2z} + O\left(\frac{1}{z^2}\right). \tag{2.57}$$

In particular, this gives

$$\psi(1+d-u) = \log(1+d-u) - \frac{1}{2(d+1)} + O\left(\frac{1}{(d+1)(d+2)}\right)$$

and

$$\psi(n+3-2u) = \log(n+2-2u) + \frac{1}{2(n+2-2u)} + O\left(\frac{1}{n^2}\right).$$

Using this in (7), we obtain

$$\begin{aligned}
&\sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d (d-u+1/2) \log(1+d-u) \\
&\qquad = \frac{1}{2} \sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d \frac{(d-u+1/2)}{d+1} \\
&\qquad\quad + O\left(\sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d \frac{(d-u+1/2)}{(d+1)(d+2)} \right) \\
&+ (2-2u)\log n - \frac{-22+40u-23u^2+4u^3}{2\,(2-u)\,(1-u)} + (-2+2u)\,\psi(5-2u) + (2-2u)\,\psi(1-u) \\
&\qquad + n\left(\log n + \frac{7-5u}{2\,(2-u)\,(1-u)} - \psi(5-2u) + \psi(1-u)\right) + O\left(\frac{1}{n}\right)
\end{aligned} \tag{2.58}$$

In the first expression on the right-hand side of (2.58) we use the trivial identity

$$\frac{d-u+1/2}{d+1} = 1 - \frac{u+1/2}{d+1},$$

to split the expression into two sums, one of which can be evaluated by means of (2.51). The other sum equals basically $-(u+1/2)$ times the sum on the left-hand side of (2.53). What is missing is the summand for $d=-1$. By (2.53), the complete sum is of the order $O(1/n)$. Using Stirling's formula it is seen that the summand for $d=-1$ is of the order $O(1/n^{1-u})$. So, combining everything, the first expression in (2.58) equals $1 + O(1/n) + O(1/n^{1-u}) = 1 + O(1/n^{1-u})$. For the second expression, we do a similar partial fraction expansion in order to apply (2.53). The result is that this second expression is of the order $O(1/n^{1-u})$. This establishes the Lemma. □

Now we are in the position to prove the Theorem.

Proof of the Theorem. Since $\lambda_{n+1-d}=\lambda_d$, an equivalent expression for the left-hand side in (2.50) is

$$-\frac{1}{2}\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d\log\lambda_d. \tag{2.59}$$

Now, we expand the logarithm according to the addition rule to obtain

$$\begin{aligned}-\frac{1}{2}\sum_{d=0}^{n+1}&\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d\\
&\cdot\log\left(\frac{1}{2^n}\frac{\Gamma(5/2-u)}{\Gamma(5/2+n/2-u)\,\Gamma(2+n/2-u)\,\Gamma(1-u)}\right)\\
&-\frac{1}{2}\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d\big(\log\Gamma(1+d-u)+\log\Gamma(2+n-d-u)\big).\end{aligned}$$

The first sum in this expression can be evaluated by means of (2.51). Therefore, we obtain for the expression on the left-hand side of (2.50)

$$\begin{aligned}n\log 2-\log\Gamma(5/2-u)+\log\Gamma(1-u)+\log\Gamma(5/2+n/2-u)&\\
+\log\Gamma(2+n/2-u)-\sum_{d=0}^{n+1}\frac{(n-2d+1)^2}{(n+1)}\binom{n+1}{d}\lambda_d\log\Gamma(1+d-u).&\end{aligned} \tag{2.60}$$

The only difficulty in obtaining the asymptotics of expression (2.60) stems from the sum. In this sum, we use Stirling's formula

$$\log\Gamma(x)=(x-1/2)\log x-x+\frac{1}{2}\log 2+\frac{1}{2}\log\pi+O\left(\frac{1}{x}\right)$$

to get

$$
\begin{aligned}
&\sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d \log \Gamma(1+d-u)\\
&= \sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d\\
&\qquad \cdot \Big((1/2+d-u) \log(1+d-u) - 1 + u - d + \frac{1}{2}\log 2 + \frac{1}{2}\log\pi + O\Big(\frac{1}{d+1}\Big)\Big)\\
&= \sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d (u - 1 + \frac{1}{2}\log 2 + \frac{1}{2}\log\pi)\\
&\qquad\quad - \sum_{d=0}^{n+1} (n-2d+1)^2 \binom{n}{d-1} + O\left(\sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)(n+2)} \binom{n+2}{d+1}\right)\\
&\qquad\quad + \sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} (1/2-u+d)\log(1-u+d). \qquad (2.61)
\end{aligned}
$$

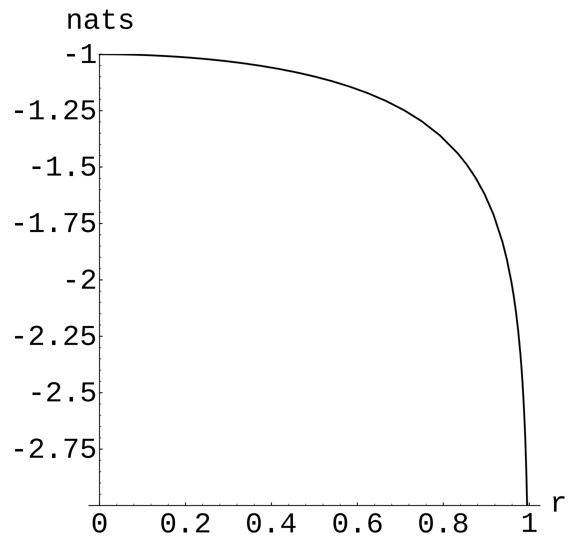
The first expression on the right-hand side of (2.61) simplifies by means of (2.51), the second by means of (2.52). For the $O(.)$ term we use (2.53). In fact, the sum on the left-hand side of (2.53) differs from the sum in the $O(.)$ term only by the summand for $d=-1$. This summand is of the order $O(1/n^{1-u})$, as is seen by Stirling's formula. Putting everything together, we obtain

$$
\begin{aligned}
&\sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d \log \Gamma(1+d-u)\\
&\qquad\qquad = 2u - 2 + \log 2 + \log \pi - (n+1) + O\left(\frac{1}{n^{1-u}}\right)\\
&\qquad\qquad + \sum_{d=0}^{n+1} \frac{(n-2d+1)^2}{(n+1)} \binom{n+1}{d} \lambda_d (1/2-u+d)\log(1-u+d). \quad (2.62)
\end{aligned}
$$

When we use this in (2.60), apply Lemma 3 to the remaining sum, and simplify, we finally arrive at (2.50). □

3. Comparison of our asymptotic redundancies for the one-parameter family $q(u)$ with those of Clarke and Barron

Let us, first, compare the formula (1.3) for the asymptotic redundancy of Clarke and Barron to that derived here (2.36) for the two-level quantum systems, in terms of the one-parameter family of probability densities $q(u)$, $-\infty < u < 1$, given in (1.10). Since the unit ball or Bloch sphere of such systems is three-dimensional in nature, we are led to set the dimension d of the parameter space in (1.3) to 3. The quantum Fisher information matrix $I(\theta)$ for that case was taken to be (1.8), while the role of the probability function $w(\theta)$ is played by $q(u)$. Under these substitutions, it was seen in the Introduction that formula (1.3) reduces to (1.12). Then, we see that for $0 \le r < 1$, the formulas (2.36) and (1.12) coincide except for the presence of the monotonically decreasing (nonclassical/quantum) term $\frac{1}{2r} \log\left(\frac{1-r}{1+r}\right)$ (see Figure 2 for a plot of this term — $\log 2 \approx .693147$ "nats" of information equalling one "bit") in (2.36). (This term would have to be replaced by -1 — that is, its limit for $r \to 0$ — to give (1.12).) In particular, the order of magnitude, $\frac{3}{2} \log n$, is precisely the same in both formulas. For the particular case $r = 0$, the asymptotic formula (2.36) (see (2.37)) precisely coincides with (1.12).



Nonclassical/quantum term ($\frac{1}{2r} \log \frac{1-r}{1+r}$) in the quantum asymptotic redundancy (2.36)

Figure 2

In the case $r = 1$, however, i.e., when we consider the boundary of the parameter space (represented by the unit sphere), the situation is slightly tricky. Due to the fact that the formula of Clarke and Barron holds only for interior points of the parameter space, we cannot expect that, in general, our formula will resemble that of Clarke and Barron. However, if the probability density, $q(u)$, is concentrated on the boundary of the sphere, then we may disregard the interior of the sphere, and may consider the boundary of the sphere as the *true* parameter space. This parameter space is *two-dimensional* and consists of interior points throughout. Indeed, the probability density $q(u)$ is concentrated on the boundary of the sphere if we choose $u = 1$ since, as we remarked in the Introduction, in the limit $u \to 1$, the distribution determined by $q(u)$ tends to the uniform distribution over the boundary of the sphere. Let us, again, (naively) attempt to apply Clarke and Barron's formula (1.3) to that case. We parameterize the boundary of the sphere by polar coordinates (ϑ, ϕ),

$$\begin{aligned} x &= \sin\vartheta\cos\varphi \\ y &= \sin\vartheta\sin\varphi \\ z &= \cos\vartheta, \\ 0 \leq \varphi \leq 2\pi,&\ 0 \leq \vartheta \leq \pi. \end{aligned}$$

The probability density induced by $q(u)$ in the limit $u \to 1$ then is $\sin\vartheta/4\pi$, the density of the uniform distribution. Using [24, eq. 8], the quantum (symmetric logarithmic derivative) Fisher information matrix turns out to be

$$\begin{pmatrix} 1 & 0 \\ 0 & \sin^2\vartheta \end{pmatrix} \quad , \tag{3.1}$$

its determinant equalling, therefore, $\sin^2\vartheta$. So, setting $d = 2$ and substituting $\sin\vartheta/4\pi$ for $w(\theta)$ and $\sin^2\vartheta$ for $I(\theta)$ in (1.3) gives $\log n + \log 2 - 1$. On the other hand, our formula (2.38), for $u = 1$, gives $\log n$. So, again, the terms differ only by a constant. In particular, the order of magnitude is again the same.

Let us now focus our attention on the asymptotic minimax redundancy (1.4) of Clarke and Barron. If in (1.4) we again set d to 3, we obtain (1.11). Clarke and Barron prove that this minimax expression is only attained by the (classical) Jeffreys' prior. In order to derive its quantum counterpart — at least, a restricted (to the family $q(u)$) version — we have to determine the behavior of

$$\min_{-\infty<u<1} \max_{0\leq r\leq 1} S(\overset{n}{\otimes}\rho, \zeta_n(u)) \tag{3.2}$$

for $n \to \infty$. We are unable to proceed in a fully rigorous manner. However, from computational data we conjecture that

$$\max_{0\leq r\leq 1} S(\overset{n}{\otimes}\rho, \zeta_n(u)) \tag{3.3}$$

is always attained at $r = 0$ (corresponding to the fully mixed state) or $r = 1$ (corresponding to a pure state). Assuming the validity of this conjecture, the maximum u_n in (3.3) is a value for which $S(\overset{n}{\otimes}\rho, \zeta_n(u))|_{r=0}$ equals $S(\overset{n}{\otimes}\rho, \zeta_n(u))|_{r=1}$. Then we are able to prove that $\lim_{n\to\infty} u_n = .5$.

Namely, by our assumption we have

$$S(\overset{n}{\otimes}\rho, \zeta_n(u_n))|_{r=0} = S(\overset{n}{\otimes}\rho, \zeta_n(u_n))|_{r=1}, \tag{3.4}$$

for any n. Let $(u_{n_k})_{k=1,2,\ldots}$ be a subsequence of the sequence (u_n) which converges to some u_0, $-\infty \leq u_0 \leq 1$. Note that we allow $u_0 = -\infty$ and $u_0 = 1$. Therefore, there is always such a subsequence. Because of (3.4) we must have

$$\lim_{k\to\infty} \frac{S(\overset{n_k}{\otimes}\rho, \zeta_{n_k}(u_{n_k}))|_{r=0}}{\log n_k} = \lim_{k\to\infty} \frac{S(\overset{n_k}{\otimes}\rho, \zeta_{n_k}(u_{n_k}))|_{r=1}}{\log n_k}. \tag{3.5}$$

By (2.37) and the fact that the error term in (2.37) is uniform in u, we know that the left-hand side in (3.5) is $3/2$. On the other hand, by (2.38) and the fact that the error term in (2.38) is uniform in u, the right-hand side in (3.5) equals $\lim_{k\to\infty}(2-u_{n_k})$. Hence, we must have $\lim_{k\to\infty} u_{n_k} = .5$. Thus, every convergent subsequence of (u_n) (including those which converge to $-\infty$ or 1, the boundary points of the interval of possible values of u_n) converges to .5. Hence, the complete sequence (u_n) converges to .5, establishing our claim. Since we have regarded $q(.5)$, that is (1.9), as the quantum counterpart of the Jeffreys' prior (because, by analogy with the classical situation, it is the normalized square root of the determinant of the quantum Fisher information matrix, $\sqrt{\det I(\theta)}$), this result could be considered to be fully parallel to that of Clarke and Barron.

We now concern ourselves with the asymptotic *maximin* redundancy. Clarke and Barron [17, 18] prove that the maximin redundancy is attained asymptotically, again, by the Jeffreys' prior. To derive the quantum counterpart of the maximin redundancy within our analytical framework, we would have to calculate

$$\max_{w} \min_{Q_n} \int_{x^2+y^2+z^2\le 1} S(\overset{n}{\otimes}\rho, Q_n)\, w(x,y,z)\, dx\, dy\, dz, \tag{3.6}$$

where Q_n varies over the $(2^{2n}-1)$-dimensional convex set of $2^n \times 2^n$ density matrices and w varies over all probability densities over the unit ball. In the classical case, due to a result of Aitchison [2, pp. 549/550], the minimum is achieved by setting Q_n to be the Bayes estimator, i.e., the average of all possible Q_n's with respect to the given probablity distribution. In the quantum domain the same assertion is true. For the sake of completeness, we include the proof in the Appendix. We can, thus, take the quantum analog of the Bayes estimator to be the Bayesian density matrix $\zeta_n(u)$. That is, we set $Q_n = \zeta_n(u)$ in (3.6). Let us, for the moment, restrict the possible w's over which the maximum is to be taken to the family $q(u)$, $-\infty < u < 1$. Thus, we consider

$$\max_{u} \int_{x^2+y^2+z^2\le 1} S(\overset{n}{\otimes}\rho, \zeta_n(u))\, q(u)\, dx\, dy\, dz. \tag{3.7}$$

By the definition (1.5) of relative entropy, we have

$$\begin{aligned} S(\overset{n}{\otimes}\rho, \zeta_n(u)) &= \operatorname{Tr}\left(\overset{n}{\otimes}\rho \log \overset{n}{\otimes}\rho\right) - \operatorname{Tr}\left(\overset{n}{\otimes}\rho \log \zeta_n(u)\right) \\ &= n\frac{(1-r)}{2}\log\frac{(1-r)}{2} + n\frac{(1+r)}{2}\log\frac{(1+r)}{2} - \operatorname{Tr}\left(\overset{n}{\otimes}\rho \log \zeta_n(u)\right), \end{aligned}$$

the second line being due to (1.7). Therefore, we get

$$\begin{aligned}\int_{x^2+y^2+z^2\le 1} & S(\overset{n}{\otimes}\rho, \zeta_n(u))\, q(u)\, dx\, dy\, dz \\ &= \left(n \int_0^1 \int_0^\pi \int_0^{2\pi} \left(\frac{(1-r)}{2} \log \frac{(1-r)}{2} + \frac{(1+r)}{2} \log \frac{(1+r)}{2} \right) \right) d\varphi\, d\vartheta\, dr \\ &\qquad\qquad - \operatorname{Tr}\big(\zeta_n(u) \log \zeta_n(u)\big) \\ &= -n \left(\frac{-7+5u}{2\,(2-u)\,(1-u)} + \psi(5-2u) - \psi(1-u) \right) + S(\zeta_n(u)). \quad (3.8)\end{aligned}$$

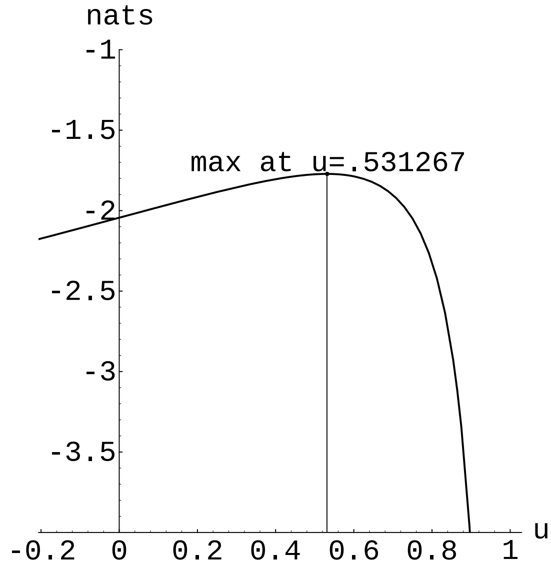
From Theorem 12, we know the asymptotics of the von Neumann entropy $S(\zeta_n(u))$. Hence, we find that the expression (3.8) is asymptotically equal to

$$\begin{aligned}\frac{3}{2} \log n &+ \left(-\frac{7}{2} + 2u \right) \log 2 \\ &- \frac{14 - 20u + 7u^2}{2\,(2-u)\,(1-u)} + \log\big(\Gamma(1-u)\big) - \log\big(\Gamma(5/2-u)\big) \\ &+ (2-2u)(\psi(5-2u) - \psi(1-u)) + O\left(\frac{1}{n^{1-u}}\right). \quad (3.9)\end{aligned}$$

We have to, first, perform the maximization required in (3.7), and then determine the asymptotics of the result. Due to the form of the asymptotics in (3.9), we can, in fact, derive the proper result by proceeding in the reverse order. That is, we first determine the asymptotics of $\int S(\overset{n}{\otimes}\rho, \zeta_n(u))\, q(u)\, dx\, dy\, dz$, which we did in (3.9), and then we maximize the u-dependent part in (3.9) with respect to u (ignoring the error term). (In Figure 3 we display this u-dependent part over the range $[-0.2, 1]$.) Of course, we do the latter step by equating the first derivative of the u-dependent part in (3.9) with respect to u to zero and solving for u. It turns out that this equation takes the appealingly simple form

$$2(1-u)^3\big(\psi'(1-u) - \psi'(5/2-u)\big) = 1. \quad (3.10)$$

Numerically, we find this equation to have the solution $u \approx .531267$, at which the asymptotic maximin redundancy assumes the value $\frac{3}{2}\log n - 1.77185 + O(1/n^{.468733})$. For $u = .5$, on the other hand, we have for the asymptotic minimax redundancy, $\frac{3}{2}\log n - 2 - \frac{1}{2}\log 2 + \frac{1}{2}\log \pi + O(1/\sqrt{n}) = \frac{3}{2}\log n - 1.77421 + O(1/\sqrt{n})$. We must, therefore, conclude that — in contrast to the classical case [17, 18] — our trial candidate ($q(.5)$) for the quantum counterpart of Jeffreys' prior can not serve as a "reference prior," in the sense introduced by Bernardo [8, 9].



u-dependent part of the asymptotic Bayes redundancy (3.9)

Figure 3

Since they are mixtures of product states, the matrices $\zeta_n(u)$ are classically — as opposed to EPR, Einstein–Podolsky–Rosen — correlated [59]. Therefore, $S(\zeta_n(u))$ must not be less than the sum of the von Neumann entropies of any set of reduced density matrices obtained from it, through computation of partial traces. For positive integers, $n_1 + n_2 + \cdots = n$, the corresponding reduced density matrices are simply $\zeta_{n_1(u)}, \zeta_{n_2(u)}, \dots$, due to the mixing [6, exercise 7.10]. Using these reduced density matrices, one can compute *conditional* density matrices and quantum entropies [13]. Clarke and Barron [17, p. 40] have an alternative expression for the redundancy in terms of conditional entropies, and it would be of interest to ascertain whether a quantum analogue of this expression exists.

Let us note that the theorem of Clarke and Barron utilized the uniform convergence property of the asymptotic expansion of the relative entropy (Kullback-Leibler divergence). Condition 2 in their paper [17] is, therefore, crucial. It assumes — as is typically the case classically — that the matrix of second derivatives, $J(\theta)$, of the relative entropy is identical to the Fisher information matrix $I(\theta)$. In the quantum domain, however, in general, $J(\theta) \geq I(\theta)$, where $J(\theta)$ is the matrix of second derivatives of the quantum relative entropy (1.5) and $I(\theta)$ is the symmetric logarithmic derivative Fisher information matrix [42, 43]. The equality holds only for special cases. For instance, $J(\theta) > I(\theta)$ does hold if $r \neq 0$ for the situation considered in this paper. The volume element of the Kubo-Mori/Bogoliubov (monotone) metric [42, 43] is given by $\sqrt{\det J(\theta)}$. This can be normalized for the two-level quantum systems to be a member ($u = 1/2$) of a one-parameter family of probability densities

$$\frac{(1-u)\,\Gamma(5/2-u)\,r\,\log\big((1+r)/(1-r)\big)\,\sin\vartheta}{\pi^{3/2}\,(3-2u)\,\Gamma(1-u)\,(1-r^2)^u}, \quad -\infty < u < 1, \tag{3.11}$$

and similarly studied, it is presumed, in the manner of the family $q(u)$ (cf. (1.10) and (2.5)) analyzed here. These two families can be seen to differ — up to the normalization factor — by the replacement of $\log\big((1+r)/(1-r)\big)$ in (3.11) by, simply, r. (These two last expressions are, of course, equal for $r = 0$.) In general, the volume element of a monotone metric over the two-level quantum systems is of the form [42, eq. 3.17]

$$\frac{r^2 \sin\vartheta}{f\big((1-r)/(1+r)\big)(1-r^2)^{1/2}(1+r)}, \tag{3.12}$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is an operator monotone function such that $f(1) = 1$ and $f(t) = tf(1/t)$. For $f(t) = (1+t)/2$, one recovers the volume element ($\sqrt{\det I(\theta)}$) of the metric of the symmetric logarithmic derivative, and for $f(t) = (t-1)/\log t$, that ($\sqrt{\det J(\theta)}$) of the Kubo-Mori/Bogoliubov metric [41, 42, 43]. (It would appear, then, that the only member of the family $q(u)$ proportional to a monotone metric is $q(.5)$, that is (1.9). The maximin result we have obtained above corresponding to $u \approx .531267$ — the solution of (3.10) — would appear unlikely, then, to extend globally beyond the family.) While $J(\theta)$ can be generated from the relative entropy (1.5) (which is a limiting case of the α-entropies [44]), $I(\theta)$ is similarly obtained from [41, eq. 3.16]

$$\operatorname{Tr} \rho_1 (\log \rho_1 - \log \rho_2)^2. \tag{3.13}$$

It might prove of interest to repeat the general line of analysis carried out in this paper, but with the use of (3.13) rather than (1.5). Also of importance might be an analysis in which the relative entropy (1.5) is retained, but the family (3.11) based on the Kubo-Mori/Bogoliubov metric is used instead of $q(u)$. Let us also indicate that if one equates the asymptotic redundancy formula of Clarke and Barron (1.3) (using $w(\theta) = q(u)$) to that derived here (2.36), neglecting the residual terms, solves for $\det(I(\theta))$, and takes the square root of the result, one obtains a prior of the form (3.12) based on the monotone function $t^{\frac{t}{1+t}}$.

As we said in the Introduction, ideally we would like to start with a (suitable well-behaved) *arbitrary* probability density on the unit ball, determine the relative entropy of $\overset{n}{\otimes}\rho$ with respect to the average of $\overset{n}{\otimes}\rho$ over the probability density, then find its asymptotics, and finally, among all such probability densities, find the one(s) for which the minimax and maximin are attained. In this regard, we wish to mention that a suitable combination of results and computations from Sec. 2 with basic facts from representation theory of $SU(2)$ (cf. [57, 10] for more information on that topic) yields the following result.

Theorem 15. *Let w be a spherically symmetric probability density on the unit ball, i.e., $w = w(x, y, z)$ depends only on $r = \sqrt{x^2 + y^2 + z^2}$. Furthermore, let $\hat{\zeta}_n(w)$ be the average $\int_{x^2+y^2+z^2\le 1} \big(\overset{n}{\otimes} \rho \big)\, w \, dx\, dy\, dz$. Then the eigenvalues of $\hat{\zeta}_n(w)$ are*

$$\lambda_d = \frac{\pi}{2^{n-1}(n-2d+1)} \int_{-1}^{1} r(1+r)^{n-d+1}(1-r)^d w(|r|)\, dr, \quad d = 0, 1, \dots, \left\lfloor \frac{n}{2} \right\rfloor, \tag{3.14}$$

with respective multiplicities

$$\frac{n-2d+1}{n+1} \binom{n+1}{d}, \tag{3.15}$$

and corresponding eigenspaces as given by (2.16).

The relative entropy of $\overset{n}{\otimes}\rho$ with respect to $\hat{\zeta}_n(w)$ is given by (2.30)*, with λ_d as given in* (3.14)*.*

We hope that this Theorem enables us to determine the asymptotics of the relative entropy and, eventually, to find, at least within the family of spherically symmetric probability densities on the unit ball, the corresponding minimax and maximin redundancies.

4. Summary

Clarke and Barron [17, 18] (cf. [45]) have derived several forms of asymptotic redundancy for arbitrarily parameterized families of probability distributions. We have been motivated to undertake this study by the possibility that their results may generalize, in some yet not fully understood fashion, to the quantum domain of noncommutative probability. (Thus, rather than probability densities, we have been concerned here with density matrices.) We have only, so far, been able to examine this possibility in a somewhat restricted manner. By this, we mean that we have limited our consideration to two-level quantum systems (rather than n-level ones, $n \ge 2$), and for the case $n = 2$, we have studied (what has proven to be) an analytically tractable one-parameter family of possible prior probability densities, $q(u)$, $-\infty < u < 1$ (rather than the totality of arbitrary probability densities). Consequently, our results can not be as definitive in nature as those of Clarke and Barron. Nevertheless, the analyses presented here indicate that our trial candidate ($q(.5)$, that is (1.9)) for the quantum counterpart of the Jeffreys' prior plays a somewhat similarly privileged — but less pronounced — role as in the classical case.

Future research might be devoted to expanding the family of probability distributions used to generate the Bayesian density matrices for $n = 2$, as well as similarly studying the n-level quantum systems ($n > 2$). (In this regard, we have examined the situation in which $n = 2^m$, and the only $n \times n$ density matrices considered are simply the tensor products of m identical 2×2 density matrices. Surprisingly, for $m = 2, 3,$

the associated trivariate candidate quantum Jeffreys' prior, taken, as throughout this study, to be proportional to the volume elements of the metrics of the symmetric logarithmic derivative (cf. [52]), have been found to be *improper* (nonnormalizable) over the Bloch sphere. The minimality of such metrics is guaranteed, however, only if "the whole state space of a spin is parameterized" [42].) In all such cases, it will be of interest to evaluate the characteristics of the relevant candidate quantum Jeffreys' prior *vis-à-vis* all other members of the family of probability distributions employed over the (n^2-1)-dimensional convex set of $n \times n$ density matrices.

We have also conducted analyses parallel to those reported above, but having, *ab initio*, set either x or y to zero in the 2×2 density matrices (1.6). This, then, places us in the realm of real — as opposed to complex (standard or conventional) quantum mechanics. (Of course, setting *both* x and y to zero would return us to a strictly classical situation, in which the results of Clarke and Barron [17, 18], as applied to binomial distributions, would be directly applicable.) Though we have — on the basis of detailed computations — developed strong conjectures as to the nature of the associated results, we have not, at this stage of our investigation, yet succeeded in formally demonstrating their validity.

In conclusion, again in analogy to classical results, we would like to raise the possibility that the quantum asymptotic redundancies derived here might prove of value in deriving formulas for the *stochastic complexity* [45, 46] (cf. [54]) — the shortest description length — of a string of n *quantum* bits. The competing possible models for the data string might be taken to be the 2×2 density matrices (ρ) corresponding to different values of r, or equivalently, different values of the von Neumann entropy, $S(\rho)$.

Appendix: The quantum Bayes estimator achieves the minimum average entropy

Let P_θ, $\theta \in \Theta$, be a family of density matrices, and let $w(\theta)$, $\theta \in \Theta$, be a family of probability distributions.

Theorem 16. *The minimum*

$$\min_Q \int w(\theta) S(P_\theta, Q)\, d\theta,$$

taken over all density matrices Q, is achieved by $m = \int w(\theta) P_\theta \, d\theta$.

Proof. We look at the difference

$$\int w(\theta) S(P_\theta, Q)\, d\theta - \int w(\theta) S(P_\theta, m)\, d\theta,$$

and show that it is nonnegative. Indeed,

$$\begin{aligned}
\int w(\theta)S(P_\theta,Q)\,d\theta &- \int w(\theta)S(P_\theta,m)\,d\theta \\
&= \int w(\theta)\,\mathrm{Tr}(P_\theta\log P_\theta - P_\theta\log Q)\,d\theta - \int w(\theta)\,\mathrm{Tr}(P_\theta\log P_\theta - P_\theta\log m)\,d\theta \\
&= \int w(\theta)\,\mathrm{Tr}\big(P_\theta(\log m - \log Q)\big)\,d\theta \\
&= \mathrm{Tr}\left(\Big(\int w(\theta)P_\theta\,d\theta\Big)(\log m - \log Q)\right) \\
&= Tr\big(m(\log m - \log Q)\big) \\
&= S(m,Q) \geq 0,
\end{aligned}$$

since relative entropies are nonnegative [39]. □

Acknowledgments

Christian Krattenthaler did part of this research at the Mathematical Sciences Research Institute, Berkeley, during the Combinatorics Program 1996/97. Paul Slater would like to express appreciation to the Institute for Theoretical Physics for computational support. This research was undertaken, in part, to respond to concerns (regarding the rationale for the presumed quantum Jeffreys' prior) conveyed to him by Walter Kohn and members of the informal seminar group he leads. The co-authors are grateful to Helmut Prodinger and Peter Grabner for their hints regarding the asymptotic computations, to Ira Gessel for bringing them into initial contact *via* the Internet, and to A. R. Bishop and an anonymous referee of [52].

References

[1] S. L. Adler, *Quaternionic Quantum Mechanics and Quantum Fields*. Oxford: New York, 1995.
[2] J. Aitchison, "Goodness of prediction fit," *Biometrika*, vol. 62, no. 3, pp.547-554, 1975.
[3] A. Bach and A. Srivastav, "A characterization of the classical states of the quantum harmonic oscillator by means of de Finetti's theorem" *Comm. Math. Phys.*, vol. 123, no. 3, pp. 453-462, 1989.
[4] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello, *Stabilisation of Quantum Computations by Symmetrisation*, Los Alamos preprint archive, quant-ph/9604028, 25 Apr. 1996.
[5] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, "General Fidelity Limit for Quantum Channels," *Phys. Rev. A*, vol. 54, no. 6, pp. 4707-4711, Dec 1996.
[6] E. G. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics*, Addison-Wesley: Reading, 1981.
[7] C. H. Bennett, "Quantum information and computation," *Physics Today*, vol. 48, no. 10, pp. 24-30, Oct. 1995.

[8] J. M. Bernardo, "Reference posterior distributions for Bayesian inference,", *J. Roy. Statist. Soc. B*, vol. 41, pp. 113-147, 1979.

[9] J. M. Bernardo and A. F. M. Smith, *Bayesian theory.* Wiley: New York, 1994.

[10] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics*, Addison–Wesley: Massachusetts, 1981.

[11] S. L. Braunstein and G. J. Milburn, "Dynamics of statistical distance: quantum limits of two-level clocks," *Phys. Rev. A*, vol. 51, no. 3, pp. 1820-1826, Mar. 1995.

[12] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Phys. Rev. A*, vol. 54, no. 2, pp. 1098-1105, Aug. 1996.

[13] N. J. Cerf and C. Adami, *Quantum theory of entanglement*, Los Alamos preprint archive, quant-ph/9605039, 28 May 1996.

[14] N. N. Chentsov, *Statistical Decision Rules and Optimal Inference*. Amer. Math. Soc.: Providence, 1982.

[15] B. S. Clarke, "Implications of reference priors for prior information and for sample size," *J. Amer. Statist. Assoc.*, vol. 91, no. 433, pp. 173-184, March 1996.

[16] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inform. Theory*, vol. 36, no. 3, pp. 453-471, May, 1990.

[17] B. S. Clarke and A. R. Barron, "Jeffreys' prior is asymptotically least favorable under entropy risk," *J. Statist. Planning and Inference*, vol. 41, no. 1, pp. 37-61, Aug. 1994.

[18] B. S. Clarke and A. R. Barron, "Jeffreys' prior yields the asymptotic minimax redundancy," in *IEEE-IMS Workshop on Information Theory and Statistics*, Piscataway, NJ: IEEE, 1995, p. 14.

[19] R. Cleve and D. P. DiVincenzo, "Schumacher's quantum data compression as a quantum computation," *Phys. Rev. A*, vol. 54, no. 4, pp. 2636-2650, Oct. 1996.

[20] I. Csiszár, "Universal Compression and Retrieval," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 862-863, May, 1995.

[21] S. De Leo and P. Rotelli, "Odd-dimensional translation between complex and quaternionic quantum mechanics," *Progress Theor. Phys.*, vol. 96, no. 1, pp. 247-255, July 1996.

[22] D. I. Fivel, "How interference effects in mixtures determine the rules of quantum mechanics," *Phys. Rev. A*, vol. 50, no. 3, pp. 2108-2119, Sept. 1994.

[23] B. R. Frieden and B. H. Soffer, "Lagrangians of physics and the game of Fisher-information transfer," *Phys. Rev. E*, vol. 52, no. 3, pp. 2274-2286, Sept. 1995.

[24] A. Fujiwara and H. Nagaoka, "Quantum Fisher metric and estimation for pure state models, *Phys. Lett. A*, vol. 201, pp. 119-124, 1995.

[25] G. Gasper and M. Rahman, *Basic Hypergeometric Series*, Encyclopedia of Mathematics And Its Applications 35, Cambridge University Press, Cambridge, 1990.

[26] I. J. Good, *Math. Rev.*, 95k:62011, Nov. 1995.

[27] I. J. Good, "Utility of a distribution'," *Nature*, vol. 219, no. 5161, p. 1392, 28 Sept. 1968.

[28] I. J. Good, "What is the use of a distribution," in *Multivariate Analysis-II* (P. R. Krishnaiah, Ed.). New York: Academic Press, 1969, pp. 183-203.

[29] K. R. W. Jones, "Principles of quantum inference," *Ann. Phys. (NY)*, vol. 207, no. 1, pp. 140-170, 1991.

[30] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Mod. Opt.* , vol. 41, no. 12, pp. 2343-2349, 1994.

[31] R. L. Kashyap, "Prior probability and uncertainty," *IEEE Trans. Inf. Th.*, vol. 17, no. 6, pp. 641-650, Nov. 1971.

[32] R. E. Kass and L. Wasserman, "The selection of prior distributions by formal rules," *J. Amer. Statist. Assoc.*, vol. 91, no. 435, pp. 1343-1370, Sept. 1995.

[33] R. Kirchevsky, *Universal Compression and Retrieval.* Dordrecht: Kluwer, 1994.

[34] E. G. Larson and P. R. Dukes, "The evolution of our probability image for the spin orientation of a spin-1/2–ensemble — connection with information theory and Bayesian statistics," in *Maximum Entropy and Bayesian Methods* (W. T. Grandy, Jr. and L. H. Schick, Eds.). Dordrecht: Kluwer, 1991, pp. 181-189.

[35] H.-K. Lo, "Quantum coding theorem for mixed states," *Opt. Commun.*, vol. 119, pp. 552-556, Sept. 1995.

[36] S. Massar and S. Popescu, "Optimal extraction of information from finite quantum ensembles," *Phys. Rev. Lett.*, vol. 74, no. 8, pp. 1259-1263, Feb. 1995.

[37] T. Matsushima, H. Inazumi, and S. Hirasawa, "A class of distortionless codes designed by Bayes decision theory," *IEEE Trans. Inf. Th.*, vol. 37, no. 5, pp. 1288-1293, Sept. 1991.

[38] S. G. Mohanty, *Lattice Path Counting and Applications*, Academic Press, New York, 1979.

[39] M. Ohya and D. Petz, *Quantum Entropy and Its Use.* Berlin: Springer-Verlag, 1993.

[40] A. Peres, *Quantum Theory: Concepts and Methods.* Dordrecht: Kluwer, 1993.

[41] D. Petz, "Geometry of canonical correlation on the state space of a quantum system," *J. Math. Phys.*, vol. 35, pp. 780-795, Feb. 1994.

[42] D. Petz and C. Sudar, "Geometries of quantum states," *J. Math. Phys.*, vol. 37, pp. 2662-2673, June 1996.

[43] D. Petz and G. Toth, "The Bogoliubov inner product in quantum statistics," *Lett. Math. Phys.*, vol. 27, pp. 205-216, 1993.

[44] D. Petz and H. Hasegawa, "On the Riemannian metric of α-entropies of density matrices," *Lett. Math. Phys.*, vol. 38, pp. 221-225, 1996.

[45] J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 40-47, Jan. 1996.

[46] J. Rissanen, *Stochastic Complexity in Statistical Inquiry.* World Scientific: Singapore, 1989.

[47] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, no. 4, pp. 2738-2747, April 1995.

[48] L. J. Slater, *Generalized Hypergeometric Functions*, Cambridge University Press, Cambridge, 1966.

[49] P. B. Slater, "Applications of quantum and classical Fisher information to two-level complex and quaternionic and three-level complex systems," *J. Math. Phys.*, vol. 37, no. 6, pp. 2682-2693, June 1996.

[50] P. B. Slater, "Quantum Fisher-Bures information of two-level systems and a three-level extension," *J. Phys. A*, vol. 29, pp. L271-L275, 21 May 1996.

[51] P. B. Slater, "The quantum Jeffreys' prior/Bures metric volume element for squeezed thermal states and a universal coding conjecture," *J. Phys. A* (to appear).

[52] P. B. Slater, *Universal Coding of Multiple Copies of Two-Level Quantum systems*, March 1996.

[53] N. J. A. Sloane and S. Plouffe, *The Encyclopaedia of Integer Sequences*, Academic Press, San Diego, 1995.

[54] K. Svozil, *Quantum algorithmic information theory*, Los Alamos preprint archive, quant-ph/9510005, 5 Oct. 1995.

[55] S. J. Szarek and D. Voiculescu, "Volumes of restricted Minkowski sums and the free analogues of the entropy power inequality," *Commun. Math. Phys.*, vol. 178, no. 3, pp. 563-570, 1996.

[56] X. Viennot, *Une Théorie Combinatoire des Polynômes Orthogonaux Generaux*, UQAM: Montreal, Quebec, 1983.

[57] N. J. Vilenkin and A. U. Klimyk, *Representation of Lie Groups and Special Functions*, vol. 1, Kluwer: Dordrecht, Boston, London, 1991.

[58] A. Wehrl, "General properties of entropy," *Rev. Mod. Phys.*, vol. 50, no. 2, pp. 221-260, Apr. 1978.

[59] R. F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model," *Phys. Rev. A*, vol. 40, no. 8, pp. 4277-4281, 15 Oct. 1989.

Christian Krattenthaler Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Vienna, Austria

E-mail address: kratt@pap.univie.ac.at

WWW: http://radon.mat.univie.ac.at/People/kratt

Paul B. Slater, Community and Organization Research Institute, University of California, Santa Barbara, CA 93106-2150

E-mail address: slater@itp.ucsb.edu