Supplementary material for “Exact Bayesian evidence for mixture models: from two components to infinity”

Matthew D. Schwartz[*]
Department of Physics, Harvard University

Subhabrata Sen
Department of Statistics, Harvard University

September 28, 2026

This supplement contains the proof details, numerical methods, and additional applications for the main paper. Section S1 gives the details of the ridge computation, with the proof of Theorem 6 of the main paper, and the localization proof for the Gaussian location mixture of Section 5.3 of the main paper; Section S2 the boundary-stratum computations with their numerical evaluation; and Section S3 the numerical methods, the numerical checks for the main-text computations, and the holonomic-structure results. Section S4 treats the proportional regime in which the number of states grows with the sample. Sections S5–S10 are the six further applications summarized in Section 6.4 of the main paper: capture–recapture, the Roman die study, the environmental-DNA panel, record linkage, the partition-prior comparison, and mutational signatures. Numbered results of the main paper are cited as “Theorem 7 of the main paper”; results numbered S1, S2, . . . belong to this supplement.

S1 Details of the ridge computation

We give here the computations underlying Theorems 6 and 7 of the main paper. Proposition S1 applies the allocation-sum assembly to the binary model, where the result can be compared with the closed form of Theorem 1. Section S1.2 proves Theorem 6. The proof contains the step covariance, the Stirling prefactors, the symbolic transverse expansion, and the endpoint estimates, and Remark S1 lists the changes needed for Theorem 7. Section S1.3 proves the analogous localization on the population ray for the Gaussian location mixture (Proposition 1 in Section 5.3 of the main paper).

S1.1 The binary model

For the binary model at symmetric counts the allocation-sum collapse of Section 5.1 of the main paper produces a two-variable exponent $E_2(x, y)$ on the allocation domain, and the exponent is flat in every direction: E_2 is constant, equal to $-2\log 2$, on the whole two-dimensional domain.

[*]Work done with Anthropic, but results and views are not endorsed by Anthropic.

This constancy is the allocation-sum image of the fact that at the symmetric data direction every mixture parameter with $\sigma_0\theta + \sigma_1\rho = 1/2$ is a maximum-likelihood point. The power of n then comes entirely from the prefactors of the local central limit theorem and Stirling's formula, which contribute $n^{-5/2}$ pointwise on an allocation lattice of n^2 points; the net result is $Z \cdot 4^n \sim n^{-1/2}$, i.e. $\lambda = 1/2$, matching the expansion of the closed form. The constant in that expansion is a sharper test of the allocation-sum assembly than the learning coefficient, which follows from the powers of n alone, and the following proposition computes the constant from the allocation sum.

Proposition S1. *For the binary model at symmetric counts, the allocation-sum computation gives*

$$Z(n,n)\,4^n \;=\; n^{-1/2} \iint_D \hat{p}_2(x,y)\,dx\,dy \;\cdot\, (1+o(1)), \qquad \hat{p}_2(x,y) \;=\; \frac{\sqrt{\pi}}{2\,x(2-x)}\,, \tag{S1.1}$$

where D is the allocation domain, and the y-dependence of the prefactor cancels between the local-CLT covariance and the Stirling factors of the weight. The domain integral evaluates to

$$\iint_D \hat{p}_2 \;=\; \frac{\sqrt{\pi}}{2} \int_0^2 \frac{1-|1-x|}{x(2-x)}\,dx \;=\; \sqrt{\pi}\,\log 2\,, \tag{S1.2}$$

in agreement with the constant obtained from Theorem 1 of the main paper.

The agreement with Theorem 1 of the main paper is a simultaneous check of the three ingredients used in the binomial-mixture assembly: the local CLT normalization, the Stirling prefactors, and the flat-locus integration.

S1.2 Proof of Theorem 6

The starting point is the collapsed allocation sum, Eq. (5.1) in Section 5.1 of the main paper. The proof has four parts, which are the four localization estimates listed after the statement of Theorem 6 in the main paper. A non-asymptotic bound, valid at every lattice point, controls the sum away from the ridge (Lemma S1). A local limit theorem for the tilted allocation walk, uniform over compact sets of tilts, gives the pointwise prefactor of ϕ_n together with the step covariance (Lemma S2). Stirling's formula for the three Beta brackets and a uniform version of the quartic expansion of Lemma 2 of the main paper then give the transverse sum at fixed ridge coordinate (Lemmas S3 and S4, Proposition S2). Finally, direct bounds on the rows near the two ends of the ridge, where the tilt degenerates, allow the ridge integral to be completed to $[0,16]$ (Lemma S6). Throughout this subsection K denotes a compact subset of $(0,16)$, with $x_- = \min K$ and $x_+ = \max K$, so that $K \subseteq [x_-, x_+] \subset (0,16)$. None of the constants below is optimized. Constants written M_K, M'_K and n_K depend only on K and may change from line to line.

Setting. The data are $U_v = n\,c_v$ with $c_v = \binom{4}{v}$, $v \in \{0,\dots,4\}$, so that $N = 16n$ and

$$\sum_v c_v = 16\,, \qquad \sum_v v\,c_v = \sum_v (4-v)\,c_v = 32\,, \qquad \sum_v v^2 c_v = 80\,, \qquad c_v = c_{4-v}\,. \tag{S1.3}$$

Write $f_v(t) = t^{4-v}(1-t)^v$, so that $c_v f_v(t)$ is the $\mathrm{Bin}(4, 1-t)$ probability of v and $\sum_v c_v f_v(t) = 1$. As in Section 3.1 of the main paper, $D(a,b) = a!\,b!/(a+b+1)! = \int_0^1 t^a(1-t)^b\,dt$ for integers $a,b \geq 0$. The weight of Section 5.1 of the main paper is

$$w(m,\ell,n) = D(m, 16n-m)\; D(4m-\ell,\, \ell)\; D(32n-4m+\ell,\, 32n-\ell)\,, \tag{S1.4}$$

and $\phi_n(m,\ell)=[\zeta^m\eta^\ell]\,G(\zeta,\eta)^n$ with $G(\zeta,\eta)=\prod_v(1+\zeta\eta^v)^{c_v}$. The allocation lattice is the set $\mathcal{L}_n$ of pairs (m,ℓ) with $\phi_n(m,\ell)\neq 0$. On it $0\le m\le 16n$, $0\le \ell\le 4m$ and $0\le 32n-\ell\le 4(16n-m)$, so all six arguments in Eq. (S1.4) are nonnegative, and $\mathcal{L}_n$ has at most $(16n+1)(32n+1)$ points. We put

$$S_n(m,\ell)=2^{64n}\,\phi_n(m,\ell)\,w(m,\ell,n)\,,\qquad Z(n)\,2^{64n}=\sum_{(m,\ell)\in\mathcal{L}_n}S_n(m,\ell)\,,\tag{S1.5}$$

and use the coordinates $x=m/n$, $y=\ell/n$, the transverse offset $j=\ell-2m\in\mathbb{Z}$ and $\epsilon=j/n=y-2x$. We put $E^*=-64\log 2$, the ridge value of Lemma 1 of the main paper. For a lattice row we write $\mathrm{row}_n(m)=\sum_\ell S_n(m,\ell)$. For probability vectors P,Q on $\{0,\dots,4\}$, $D(Q\|P)=\sum_v Q_v\log(Q_v/P_v)$ and $\|P-Q\|_1=\sum_v|P_v-Q_v|$, and we use Pinsker's inequality $D(Q\|P)\ge\frac12\|P-Q\|_1^2$. Throughout, $Q=\mathrm{Bin}(4,\frac12)$, that is $Q_v=c_v/16$, and for $(\sigma,\theta,\rho)\in[0,1]^3$

$$p_v(\sigma,\theta,\rho)=\sigma f_v(\theta)+(1-\sigma)f_v(\rho)\,,\qquad P_v(\sigma,\theta,\rho)=c_v\,p_v(\sigma,\theta,\rho)\,.\tag{S1.6}$$

These are the cell values and the reduced-state law of the mixture. Since $16p_v=P_v/Q_v$,

$$2^{64n}\prod_v p_v^{nc_v}=\exp\Big\{n\sum_v c_v\log(16p_v)\Big\}=e^{-16n\,D(Q\|P)}\,.\tag{S1.7}$$

The proof uses two exact identities. The first is the binomial expansion behind the allocation sum of Section 3.1 of the main paper, written before integration: for all $(\sigma,\theta,\rho)\in[0,1]^3$,

$$\prod_v p_v(\sigma,\theta,\rho)^{nc_v}=\sum_{(m,\ell)\in\mathcal{L}_n}\phi_n(m,\ell)\,\sigma^m(1-\sigma)^{16n-m}\,\theta^{4m-\ell}(1-\theta)^\ell\,\rho^{32n-4m+\ell}(1-\rho)^{32n-\ell}\,.\tag{S1.8}$$

Indeed, expanding each factor $p_v^{U_v}$ binomially, the term with x_v copies of $\sigma f_v(\theta)$ is $\binom{U_v}{x_v}\sigma^{x_v}(1-\sigma)^{U_v-x_v}\theta^{(4-v)x_v}(1-\theta)^{vx_v}\rho^{(4-v)(U_v-x_v)}(1-\rho)^{v(U_v-x_v)}$. Multiplying over v and grouping by $m=|x|$ and $\ell=\sum_v vx_v$, the exponents are $\sum_v(4-v)x_v=4m-\ell$, $\sum_v(4-v)(U_v-x_v)=32n-4m+\ell$ and $\sum_v v(U_v-x_v)=32n-\ell$ by Eq. (S1.3), and the weighted count of allocations at fixed (m,ℓ) is $\phi_n(m,\ell)$. Integrating Eq. (S1.8) over the unit cube recovers the collapsed sum $Z(n)=\sum\phi_n w$. The second identity is the random-walk representation of ϕ_n. For $\zeta,\eta>0$ let $\pi_v=\pi_v(\zeta,\eta)=\zeta\eta^v/(1+\zeta\eta^v)$ and let $\mathbf{X}_n$ be the sum of $16n$ independent steps in $\mathbb{Z}^2$, nc_v of them of type v, a step of type v being $(1,v)$ with probability π_v and $(0,0)$ otherwise. Expanding $G(\zeta,\eta)^n=\prod_v(1+\zeta\eta^v)^{nc_v}$ gives, for all $(m,\ell)\in\mathbb{Z}^2$,

$$\begin{gathered}\mathbb{P}_{\zeta,\eta}(\mathbf{X}_n=(m,\ell))=\frac{\phi_n(m,\ell)\,\zeta^m\eta^\ell}{G(\zeta,\eta)^n}\,,\\ \mu(\zeta,\eta)=\sum_v c_v\pi_v\begin{pmatrix}1\\ v\end{pmatrix},\qquad \Sigma(\zeta,\eta)=\sum_v c_v\pi_v(1-\pi_v)\begin{pmatrix}1\\ v\end{pmatrix}\begin{pmatrix}1\\ v\end{pmatrix}^\top,\end{gathered}\tag{S1.9}$$

where $n\mu$ and $n\Sigma$ are the mean and covariance of $\mathbf{X}_n$. Since the vectors $(1,v)$ span $\mathbb{R}^2$ and every $\pi_v(1-\pi_v)>0$, $\Sigma(\zeta,\eta)$ is positive definite. With $\mathsf{M}=\sum_v c_v\binom{1}{v}\binom{1}{v}^\top=\left(\begin{smallmatrix}16&32\\32&80\end{smallmatrix}\right)$ by Eq. (S1.3), whose eigenvalues are $48\pm32\sqrt2$, we have $\Sigma\succeq\min_v\pi_v(1-\pi_v)\,\mathsf{M}$ and we put $\lambda_0=48-32\sqrt2>0$.

Lemma S1 (Off-ridge domination)**.** *For every $n\ge1$ and every $(m,\ell)\in\mathcal{L}_n$, with $\sigma^*=m/(16n)$, $\theta^*=(4m-\ell)/(4m)$ and $\rho^*=(32n-4m+\ell)/(4(16n-m))$,*

$$S_n(m,\ell)\;\le\;e^{-16n\,D(Q\|P^*)}\;\le\;1\,,\qquad P^*=P(\sigma^*,\theta^*,\rho^*)\,,\tag{S1.10}$$

where for $m=0$ (which forces $\ell=0$) θ^* is arbitrary and for $m=16n$ (which forces $\ell=32n$) ρ^* is arbitrary. Moreover, for $0<m<16n$ and $j=\ell-2m$,

$$S_n(m,\ell) \;\le\; \exp\Big\{-\frac{n\,j^4}{32\,m^2(16n-m)^2}\Big\} \;\le\; \exp\Big\{-\frac{j^4}{131072\,n^3}\Big\}, \tag{S1.11}$$

and $j=0$ when $m\in\{0,16n\}$. Consequently, for every $\delta>0$,

$$\sum_{(m,\ell)\in\mathcal{L}_n:\,|\ell-2m|\ge n^{3/4+\delta}} S_n(m,\ell) \;\le\; (16n+1)(32n+1)\,e^{-n^{4\delta}/131072}. \tag{S1.12}$$

Proof. For integers $a,b\ge 0$, $D(a,b)=\int_0^1 t^a(1-t)^b dt \le \sup_{[0,1]} t^a(1-t)^b = a^a b^b/(a+b)^{a+b}$, with $0^0=1$. Applying this to the three factors of Eq. (S1.4), whose maximizers are $t=\sigma^*,\ \theta^*,\ \rho^*$ respectively,

$$w(m,\ell,n) \;\le\; (\sigma^*)^m(1-\sigma^*)^{16n-m}\,(\theta^*)^{4m-\ell}(1-\theta^*)^{\ell}\,(\rho^*)^{32n-4m+\ell}(1-\rho^*)^{32n-\ell}. \tag{S1.13}$$

The right side of Eq. (S1.13) is positive with the convention $0^0=1$, and it is exactly the monomial multiplying $\phi_n(m,\ell)$ in Eq. (S1.8) at $(\sigma^*,\theta^*,\rho^*)$. Since every term of Eq. (S1.8) is nonnegative, $\phi_n(m,\ell)$ times this monomial is at most $\prod_v p_v(\sigma^*,\theta^*,\rho^*)^{nc_v}$. Multiplying by Eq. (S1.13) and by 2^{64n} and using Eq. (S1.7) gives Eq. (S1.10).

For Eq. (S1.11) we compare second factorial moments, using $0\le v(v-1)\le 12$ on $\{0,\dots,4\}$. Since $c_v f_v(t)$ is the $\mathrm{Bin}(4,1-t)$ law, $\sum_v v(v-1)\,P_v(\sigma,\theta,\rho)=12\big[\sigma(1-\theta)^2+(1-\sigma)(1-\rho)^2\big]$ and $\sum_v v(v-1)\,Q_v=3$. At the matched point $1-\theta^*=\frac12+\frac{j}{4m}$ and $1-\rho^*=\frac12-\frac{j}{4(16n-m)}$, and

$$\sigma^*\Big(\frac12+\frac{j}{4m}\Big)^2+(1-\sigma^*)\Big(\frac12-\frac{j}{4(16n-m)}\Big)^2-\frac14=\frac{j^2}{256\,n}\Big(\frac1m+\frac{1}{16n-m}\Big)=\frac{j^2}{16\,m(16n-m)}, \tag{S1.14}$$

because the linear terms $\sigma^* j/(4m)=j/(64n)$ and $-(1-\sigma^*)j/(4(16n-m))=-j/(64n)$ cancel. Hence $\big|\sum_v v(v-1)(P_v^*-Q_v)\big|=\frac{3j^2}{4m(16n-m)}$, so $\|P^*-Q\|_1\ge\frac{j^2}{16m(16n-m)}$, and Pinsker's inequality gives $16nD(Q\|P^*)\ge\frac{8nj^4}{256m^2(16n-m)^2}$, which is the first bound in Eq. (S1.11). The second follows from $m(16n-m)\le 64n^2$. If $m=0$ then $\ell=0$, and if $m=16n$ then $\ell=32n$, so $j=0$ in both cases. Finally Eq. (S1.12) follows from Eq. (S1.11) and the count of lattice points, since the rows $m\in\{0,16n\}$, where Eq. (S1.11) is not asserted, consist of the single points $j=0$ and do not meet the region $|j|\ge n^{3/4+\delta}$. □

The bound Eq. (S1.10) has a continuum counterpart. Each bracket rate $\varrho(\alpha,\beta)=\alpha\log\frac{\alpha}{\alpha+\beta}+\beta\log\frac{\beta}{\alpha+\beta}$ of γ in Eq. (5.3) of the main paper (see Eq. (S1.25) and Eq. (S1.26) below) is the logarithm of the corresponding factor of Eq. (S1.13) per unit n. Also $s(x,y)\le\log G(\hat\zeta,\hat\eta)-x\log\hat\zeta-y\log\hat\eta$ at the tilt $(\hat\zeta,\hat\eta)$ with $\hat\zeta\hat\eta^v=\sigma^* f_v(\theta^*)/((1-\sigma^*)f_v(\rho^*))$, for which $G(\hat\zeta,\hat\eta)=\prod_v p_v^{c_v}/((1-\sigma^*)^{16}\rho^{*32}(1-\rho^*)^{32})$. The same cancellation of monomials then gives, for all (x,y) in the interior of the allocation domain,

$$E(x,y)-E^* \;\le\; -16\,D\big(Q\,\big\|\,P\big(\tfrac{x}{16},\tfrac{4x-y}{4x},\tfrac{32-4x+y}{4(16-x)}\big)\big) \;\le\; -\frac{(y-2x)^4}{131072}, \tag{S1.15}$$

and the bound extends to the boundary points with $0<x<16$ by continuity of s, γ and the divergence as functions of (x,y). Hence the ridge $y=2x$ is the whole maximal set of E, with equality in the first inequality of Eq. (S1.15) exactly on the ridge, where $P=Q$.

Lemma S2 (Local limit theorem for the tilted walk)**.** *For every compact set $T \subset (0,\infty)^2$ there is $C_T < \infty$ such that for all $n \geq 1$, all $(\zeta,\eta) \in T$ and all $\mathbf{z} \in \mathbb{Z}^2$, with $\mu = \mu(\zeta,\eta)$ and $\Sigma = \Sigma(\zeta,\eta)$ as in Eq.* (S1.9),

$$\Big|\mathbb{P}_{\zeta,\eta}(\mathbf{X}_n = \mathbf{z}) - \frac{1}{2\pi n\sqrt{\det\Sigma}}\exp\Big\{-\frac{(\mathbf{z}-n\mu)^\top\Sigma^{-1}(\mathbf{z}-n\mu)}{2n}\Big\}\Big| \;\leq\; C_T\, n^{-3/2}\,. \tag{S1.16}$$

At the ridge tilt $(\zeta,\eta) = (\zeta^*,1)$, $\zeta^* = x/(16-x)$, $0 < x < 16$, every step type has the same probability, $\pi_v = \varpi$ for all v with $\varpi = x/16$, and

$$\mu = (x, 2x)\,, \qquad \Sigma = \varpi(1-\varpi)\begin{pmatrix} 16 & 32 \\ 32 & 80 \end{pmatrix}, \qquad \sqrt{\det\Sigma} = 16\varpi(1-\varpi) = \frac{x(16-x)}{16}\,, \tag{S1.17}$$

and $(\Sigma^{-1})_{22} = 16/(x(16-x))$. Hence for every compact $K \subset (0,16)$ there is M_K such that, for all $n \geq 1$ and all integers m,ℓ with $x = m/n \in K$,

$$\phi_n(m,\ell) = \frac{(1+\zeta^*)^{16n}}{(\zeta^*)^m}\Big[\frac{16}{2\pi n\, x(16-x)}\exp\Big\{-\frac{8(\ell-2m)^2}{n\, x(16-x)}\Big\} + r_n(m,\ell)\Big], \qquad |r_n(m,\ell)| \leq M_K n^{-3/2}\,. \tag{S1.18}$$

Proof. Let $\delta_T = \min_T \min_v \min(\pi_v, 1-\pi_v) > 0$ and $p_T = \delta_T(1-\delta_T)$, so that $\pi_v(1-\pi_v) \geq p_T$ and $\Sigma \succeq p_T \mathsf{M} \succeq p_T\lambda_0 I$ on T. For $\vartheta = (\vartheta_1,\vartheta_2) \in \mathbb{R}^2$ write $a_v = \vartheta_1 + v\vartheta_2$, so that $|a_v| \leq \sqrt{17}\,|\vartheta|$ and $\sum_v c_v a_v^2 = \vartheta^\top\mathsf{M}\vartheta$. The characteristic function of one step of type v is $\chi_v(a_v)$ with $\chi_v(a) = 1 - \pi_v + \pi_v e^{ia}$, and Fourier inversion on $\mathbb{Z}^2$ gives

$$\mathbb{P}_{\zeta,\eta}(\mathbf{X}_n = \mathbf{z}) = \frac{1}{(2\pi)^2}\int_{[-\pi,\pi]^2}\Psi_n(\vartheta)\, e^{-i\mathbf{z}\cdot\vartheta}\, d\vartheta\,, \qquad \Psi_n(\vartheta) = \prod_v \chi_v(a_v)^{nc_v}\,. \tag{S1.19}$$

Since $|\chi_v(a)|^2 = 1 - 2\pi_v(1-\pi_v)(1-\cos a) \leq \exp\{-2\pi_v(1-\pi_v)(1-\cos a)\}$,

$$|\Psi_n(\vartheta)| \leq e^{-np_T\Upsilon(\vartheta)}\,, \qquad \Upsilon(\vartheta) = \sum_v c_v(1-\cos(\vartheta_1 + v\vartheta_2))\,. \tag{S1.20}$$

On the torus $[-\pi,\pi]^2$ the function Υ vanishes only at $\vartheta = 0$: the term $v = 0$ forces $\vartheta_1 = 0$ and then the term $v = 1$ forces $\vartheta_2 = 0$. This is the statement that the step lattice is $\mathbb{Z}^2$.

Central region. Let $r_0 = 1/(8\sqrt{17})$. For $|\vartheta| \leq r_0$ all $|a_v| \leq \frac{1}{8}$, so $|\chi_v(a_v) - 1| \leq |a_v| \leq \frac{1}{8}$ and the principal logarithm $\varkappa_v = \log\chi_v$ is analytic there, with $\varkappa_v(0) = 0$, $\varkappa_v'(0) = i\pi_v$, $\varkappa_v''(0) = -\pi_v(1-\pi_v)$. From $\varkappa_v''' = \chi_v'''/\chi_v - 3\chi_v'\chi_v''/\chi_v^2 + 2(\chi_v')^3/\chi_v^3$, $|\chi_v^{(k)}| \leq 1$ and $|\chi_v| \geq \frac{7}{8}$ we get $|\varkappa_v'''| \leq 9$ for $|a| \leq \frac{1}{8}$, hence $|\varkappa_v(a) - i\pi_v a + \frac{1}{2}\pi_v(1-\pi_v)a^2| \leq \frac{3}{2}|a|^3$. The linear terms sum to $in\sum_v c_v\pi_v a_v = in\,\mu\cdot\vartheta$, so

$$\Psi_n(\vartheta)e^{-in\mu\cdot\vartheta} = e^{n\,\varkappa(\vartheta)}\,, \qquad \Big|\varkappa(\vartheta) + \tfrac{1}{2}\vartheta^\top\Sigma\,\vartheta\Big| \leq \tfrac{3}{2}\sum_v c_v|a_v|^3 \leq C_3|\vartheta|^3\,, \qquad C_3 = 24\cdot 17^{3/2}\,, \tag{S1.21}$$

for $|\vartheta| \leq r_0$. By Eq. (S1.20) and $1 - \cos a \geq 2a^2/\pi^2$ for $|a| \leq \pi$, $\operatorname{Re} n\varkappa(\vartheta) \leq -np_T\Upsilon(\vartheta) \leq -2np_T\vartheta^\top\mathsf{M}\vartheta/\pi^2 \leq -nc_T|\vartheta|^2$ with $c_T = 2p_T\lambda_0/\pi^2$, and also $-\frac{n}{2}\vartheta^\top\Sigma\vartheta \leq -\frac{n}{2}p_T\lambda_0|\vartheta|^2 \leq -nc_T|\vartheta|^2$. For complex α,β, $|e^\alpha - e^\beta| = \big|\int_0^1 e^{\beta+u(\alpha-\beta)}(\alpha-\beta)\,du\big| \leq |\alpha-\beta|\max(e^{\operatorname{Re}\alpha}, e^{\operatorname{Re}\beta})$. Therefore

$$\frac{1}{(2\pi)^2}\int_{|\vartheta|\leq r_0}\Big|e^{n\varkappa(\vartheta)} - e^{-n\vartheta^\top\Sigma\vartheta/2}\Big|\,d\vartheta \leq \frac{nC_3}{(2\pi)^2}\int_{\mathbb{R}^2}|\vartheta|^3 e^{-nc_T|\vartheta|^2}d\vartheta = \frac{3\,C_3}{16\sqrt{\pi}}\,c_T^{-5/2}\,n^{-3/2}\,. \tag{S1.22}$$

Outer region. The set $\{\vartheta \in [-\pi,\pi]^2 : |\vartheta| \geq r_0\}$ is compact and $\Upsilon > 0$ on it, so $\Upsilon_0 = \min \Upsilon$ over this set is a positive absolute constant, and by Eq. (S1.20) the contribution of this set to Eq. (S1.19) is at most $e^{-np_T\Upsilon_0}$ in absolute value. On the Gaussian side, $(2\pi)^{-2}\int_{|\vartheta|>r_0} e^{-n\vartheta^\top\Sigma\vartheta/2}d\vartheta$ is at most $(2\pi np_T\lambda_0)^{-1}e^{-np_T\lambda_0 r_0^2/2}$.

Assembly. The Gaussian Fourier integral is exact:

$$\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} e^{-n\vartheta^\top\Sigma\vartheta/2-i(\mathbf{z}-n\mu)\cdot\vartheta}d\vartheta = \frac{1}{2\pi n\sqrt{\det\Sigma}}\, e^{-(\mathbf{z}-n\mu)^\top\Sigma^{-1}(\mathbf{z}-n\mu)/(2n)}\,. \tag{S1.23}$$

Splitting both Eq. (S1.19) and Eq. (S1.23) at $|\vartheta| = r_0$, the phase $e^{-i(\mathbf{z}-n\mu)\cdot\vartheta}$ has modulus one, and the three bounds above give Eq. (S1.16) with

$$C_T = \frac{3C_3}{16\sqrt{\pi}}\, c_T^{-5/2} + \sup_{n\geq 1} n^{3/2}\Big[e^{-np_T\Upsilon_0} + \frac{e^{-np_T\lambda_0 r_0^2/2}}{2\pi np_T\lambda_0}\Big], \tag{S1.24}$$

which depends on T only through δ_T.

At $(\zeta^*,1)$ every $\pi_v = \zeta^*/(1+\zeta^*) = x/16 = \varpi$, so $\mu = \varpi\,(16,32) = (x,2x)$ and $\Sigma = \varpi(1-\varpi)\mathsf{M}$ by Eq. (S1.3), with $\det\Sigma = \varpi^2(1-\varpi)^2(16\cdot 80 - 32^2) = 256\,\varpi^2(1-\varpi)^2$ and $\Sigma^{-1} = (256\varpi(1-\varpi))^{-1}\left(\begin{smallmatrix}80 & -32\\ -32 & 16\end{smallmatrix}\right)$, which is Eq. (S1.17) since $256\varpi(1-\varpi) = x(16-x)$. For $x \in K$ the tilts $(\zeta^*(x),1)$ form a compact subset of $(0,\infty)^2$, $G(\zeta^*,1)^n = (1+\zeta^*)^{16n}$, and Eq. (S1.9) with $\mathbf{z} = (m,\ell)$, $\mathbf{z} - n\mu = (0, \ell-2m)$ gives Eq. (S1.18). $\square$

The exponent and the prefactors near the ridge. For integers $\alpha,\beta \geq 1$ the Robbins form of Stirling's formula, $k! = \sqrt{2\pi k}\, k^k e^{-k} e^{r_k}$ with $\frac{1}{12k+1} < r_k < \frac{1}{12k}$, gives

$$\begin{gathered} D(\alpha,\beta) = e^{\varrho(\alpha,\beta)}\,\frac{b_0(\alpha,\beta)}{1+(\alpha+\beta)^{-1}}\, e^r, \\ \varrho(\alpha,\beta) = \alpha\log\frac{\alpha}{\alpha+\beta} + \beta\log\frac{\beta}{\alpha+\beta}\,, \qquad b_0(\alpha,\beta) = \frac{1}{\alpha+\beta}\sqrt{\frac{2\pi\alpha\beta}{\alpha+\beta}}\,, \end{gathered} \tag{S1.25}$$

with $|r| \leq \frac{1}{12\alpha} + \frac{1}{12\beta}$. As functions of real $\alpha,\beta > 0$, both ϱ and b_0 are homogeneous: $\varrho(n\alpha,n\beta) = n\varrho(\alpha,\beta)$ and $b_0(n\alpha,n\beta) = n^{-1/2}b_0(\alpha,\beta)$. The Stirling rate γ of the weight Eq. (S1.4), in the notation of Section 5.1 of the main paper, and the corresponding prefactor are

$$\begin{aligned} \gamma(x,y) &= \varrho(x,16-x) + \varrho(4x-y,y) + \varrho(32-4x+y,32-y)\,, \\ W(x,y) &= b_0(x,16-x)\, b_0(4x-y,y)\, b_0(32-4x+y,32-y)\,. \end{aligned} \tag{S1.26}$$

For $\zeta,\eta > 0$ and real (x,y) let $\Phi(\zeta,\eta;x,y) = \log G(\zeta,\eta) - x\log\zeta - y\log\eta$, so that $s(x,y) = \inf_{\zeta,\eta>0}\Phi$ and $E = s + \gamma$ as in Section 5.1 of the main paper.

Lemma S3 (Saddle chart and pointwise asymptotics). *(a) In the variables* $(\log\zeta,\log\eta) \in \mathbb{R}^2$ *the function* $\log G$ *is strictly convex with gradient* $\mu(\zeta,\eta)$ *and Hessian* $\Sigma(\zeta,\eta)$ *of Eq.* (S1.9). *The map* $(\zeta,\eta) \mapsto \mu(\zeta,\eta)$ *is a real-analytic diffeomorphism of* $(0,\infty)^2$ *onto an open set* $\mathcal{U} \subset \mathbb{R}^2$ *containing the ridge* $\{(x,2x) : 0 < x < 16\}$*, with* $\mu(\zeta^*(x),1) = (x,2x)$*. The set* $\mathcal{U}$ *is the interior of the Newton polygon of* G*, which is the allocation domain, since the gradient of* $\log G$ *in logarithmic coordinates maps* $\mathbb{R}^2$ *onto the interior of the Newton polygon. For* $(x,y) \in \mathcal{U}$ *the* *infimum defining* $s(x,y)$ *is attained exactly at the saddle* $(\zeta,\eta)=\mu^{-1}(x,y)$*,* s *is real-analytic on* $\mathcal{U}$*, and for all integers* (m,ℓ) *with* $(m,\ell)/n=(x,y)\in\mathcal{U}$*,*

$$\phi_n(m,\ell)=e^{n\,s(x,y)}\,\mathbb{P}_{\zeta,\eta}\big(\mathbf{X}_n=(m,\ell)\big)\Big|_{(\zeta,\eta)=\mu^{-1}(x,y)}\,,\qquad n\mu(\zeta,\eta)=(m,\ell)\,. \tag{S1.27}$$

(b) Let $K\subset(0,16)$ *be compact, with* $x_-=\min K$ *and* $x_+=\max K$ *as above. There are* $\epsilon_K\in(0,\min(x_-,16-x_+)]$ *and a compact* $T_K\subset(0,\infty)^2$ *such that the closed tube* $\mathcal{T}_K=\{(x,2x+\epsilon):x\in K,\ |\epsilon|\le\epsilon_K\}$ *lies in* $\mathcal{U}$ *and* $\mu^{-1}(\mathcal{T}_K)\subseteq T_K$*. On* $\mathcal{T}_K$ *the functions* s*,* γ*,* E*,* W *and*

$$A(x,\epsilon)=\frac{W(x,2x+\epsilon)}{2\pi\sqrt{\det\Sigma\big(\mu^{-1}(x,2x+\epsilon)\big)}} \tag{S1.28}$$

are real-analytic, W *and* A *are positive, and* $E(x,2x+\epsilon)=E(x,2x-\epsilon)$*.*

(c) On the ridge $W(x,2x)=\sqrt{2}\,\pi^{3/2}/512$ *for every* $x\in(0,16)$*, and* $A(x,0)=p(x)$ *with*

$$p(x)=\frac{\sqrt{2\pi}}{64\,x(16-x)}\,, \tag{S1.29}$$

so that $\hat{p}(x)=n^{-5/2}p(x)$ *is the pointwise prefactor of Section 5.1 of the main paper.*

(d) There are M_K *and* n_K *such that for all* $n\ge n_K$ *and all* $(m,\ell)\in\mathbb{Z}^2$ *with* $x=m/n\in K$ *and* $|\epsilon|=|\ell-2m|/n\le\epsilon_K$*,*

$$\begin{aligned}S_n(m,\ell)&=n^{-5/2}A(x,\epsilon)\,e^{n[E(x,2x+\epsilon)-E^*]}\,\big(1+\xi_n(m,\ell)\big)\,,\\ |\xi_n(m,\ell)|&\le M_Kn^{-1/2}\,,\qquad\Big|\frac{A(x,\epsilon)}{A(x,0)}-1\Big|\le M_K|\epsilon|\,.\end{aligned} \tag{S1.30}$$

In particular $S_n(m,2m)=n^{-5/2}p(x)\,(1+\xi_n(m,2m))$*. Moreover* $\frac{1}{n}\log\left[\phi_n(m,\ell)\,w(m,\ell,n)\right]-E(\frac{m}{n},\frac{\ell}{n})\to0$ *as* $n\to\infty$*, uniformly over* $(m,\ell)\in\mathbb{Z}^2$ *with* $(m,\ell)/n\in\mathcal{T}_K$*, and the same holds with* $\mathcal{T}_K$ *replaced by any compact subset of* $\mathcal{U}$*. This is Eq. (5.3) of the main paper on compact subsets of* $\mathcal{U}$*.*

Proof. (a) In $(\log\zeta,\log\eta)$, $\log G=\sum_v c_v\log(1+e^{\log\zeta+v\log\eta})$ has gradient $\sum_v c_v\pi_v(1,v)=\mu$ and Hessian $\sum_v c_v\pi_v(1-\pi_v)(1,v)(1,v)^\top=\Sigma\succ0$, so it is strictly convex and its gradient map is injective. By the inverse function theorem μ is then a real-analytic diffeomorphism onto its open image $\mathcal{U}$. At $(\zeta^*,1)$ every $\pi_v=x/16$ and $\mu=(x,2x)$ by Eq. (S1.3), so the ridge lies in $\mathcal{U}$. For $(x,y)\in\mathcal{U}$, $\Phi(\cdot\,;x,y)$ is convex in $(\log\zeta,\log\eta)$ with gradient $\mu-(x,y)$, which vanishes exactly at $\mu^{-1}(x,y)$, so the infimum is attained there and only there, and $s(x,y)=\Phi(\mu^{-1}(x,y);x,y)$ is analytic. Identity Eq. (S1.27) is Eq. (S1.9) at this tilt, since $G(\zeta,\eta)^n\zeta^{-m}\eta^{-\ell}=e^{n\Phi(\zeta,\eta;x,y)}=e^{ns(x,y)}$ there.

(b) The segment $\{(x,2x):x\in K\}$ is compact in the open set $\mathcal{U}$, so a closed tube around it of some width $\epsilon_K>0$ lies in $\mathcal{U}$, and we may take $\epsilon_K\le\min(x_-,16-x_+)$. Its image T_K under the continuous map μ^{-1} is compact. On $\mathcal{T}_K$ the six arguments of ϱ and b_0 in Eq. (S1.26), namely x, $16-x$, $2x\mp\epsilon$ and $32-2x\pm\epsilon$, are at least $c_K=\min(x_-,16-x_+)>0$, so γ and W are analytic there and $W>0$. Hence $E=s+\gamma$ and A are analytic as well, since $\det\Sigma$ is positive and analytic in the tilt. For the symmetry, $c_v=c_{4-v}$ implies $G(\zeta\eta^4,\eta^{-1})=G(\zeta,\eta)$, and then by direct substitution $\Phi(\zeta\eta^4,\eta^{-1};x,4x-y)=\Phi(\zeta,\eta;x,y)$. Since $(\zeta,\eta)\mapsto(\zeta\eta^4,\eta^{-1})$ is a bijection of $(0,\infty)^2$, $s(x,4x-y)=s(x,y)$. Also $\gamma(x,4x-y)=\gamma(x,y)$ because ϱ is symmetric and $y\mapsto 4x-y$ swaps the two arguments of the second and of the third bracket in Eq. (S1.26). Hence $E(x,2x+\epsilon)=E(x,2x-\epsilon)$.

(c) On the ridge the three pairs in Eq. (S1.26) are $(x,16-x)$, $(2x,2x)$ and $(2(16-x),2(16-x))$, with b_0 equal to $\sqrt{2\pi x(16-x)/16}/16$, $\sqrt{2\pi x}/(4x)$ and $\sqrt{2\pi(16-x)}/(4(16-x))$. The product is $(2\pi)^{3/2}x(16-x)/(1024\,x(16-x))=\sqrt{2}\,\pi^{3/2}/512$, independent of x. With $\sqrt{\det\Sigma}=x(16-x)/16$ from Eq. (S1.17), $A(x,0)=\frac{16}{2\pi x(16-x)}\cdot\frac{\sqrt{2}\,\pi^{3/2}}{512}=\frac{\sqrt{2\pi}}{64\,x(16-x)}$.

(d) Let (m,ℓ) be as stated and $(\zeta,\eta)=\mu^{-1}(x,2x+\epsilon)\in T_K$. By Eq. (S1.27) and Lemma S2 at $\mathbf{z}=n\mu$, $\phi_n(m,\ell)=e^{ns}(2\pi n\sqrt{\det\Sigma})^{-1}(1+\xi')$ with $|\xi'|\le 2\pi\sqrt{\det\Sigma}\,C_{T_K}n^{-1/2}\le M_Kn^{-1/2}$, since $\det\Sigma$ is bounded on T_K. The six factorial arguments m, $16n-m$, $2m\mp j$, $32n-2m\pm j$ of Eq. (S1.4) equal n times the six arguments listed in (b), all at least $c_Kn\ge 1$ for $n\ge n_K$. Applying Eq. (S1.25) to the three brackets and using homogeneity,

$$w(m,\ell,n)=e^{\,n\gamma(x,2x+\epsilon)}\,n^{-3/2}\,W(x,2x+\epsilon)\,e^{r'},\qquad |r'|\le\frac{3}{c_Kn}+\frac{6}{12\,c_Kn}\le\frac{M_K}{n}\,,\tag{S1.31}$$

where $3/(c_Kn)$ bounds the three factors $(1+(\alpha+\beta)^{-1})^{-1}$. Multiplying, and using $2^{64n}=e^{-nE^*}$, gives the first part of Eq. (S1.30). The bound on $A(x,\epsilon)/A(x,0)-1$ holds because A is analytic and positive on the compact set $\mathcal{T}_K$. On the ridge $E(x,2x)=E^*$ by Lemma 1 of the main paper, which gives the formula for $S_n(m,2m)$. The uniform convergence follows from Eq. (S1.27), Eq. (S1.31) and Eq. (S1.16), since $n^{-1}\log\mathbb{P}(\mathbf{X}_n=n\mu)\to 0$ uniformly on T_K. □

The q_4 identity. The transverse expansion is proved symbolically for the whole family of Theorem 7 of the main paper. With a, b, c as in that theorem, the ridge is $y_{\mathrm{ridge}}=(4c/b)\,x$ and the ridge tilt is $(\zeta^*,1)$ with $\zeta^*=x/(b^4-x)$, reducing to $2x$ and $x/(16-x)$ at $\theta_0=\frac{1}{2}$. Write the saddle $(\zeta(\epsilon),\eta(\epsilon))$ of $s(x,y_{\mathrm{ridge}}+\epsilon)$ as a series in ϵ about $(\zeta^*,1)$ and use the envelope identity $\partial_\epsilon s(x,y_{\mathrm{ridge}}+\epsilon)=-\log\eta(\epsilon)$: the derivative of a Legendre value along the constraint is minus the logarithm of the conjugate variable. Then the coefficients e_2,e_3,e_4 of $\epsilon^2,\epsilon^3,\epsilon^4$ in $E(x,y_{\mathrm{ridge}}+\epsilon)$ require only the first three orders of the saddle expansion. Solving those orders in closed form and adding the y-derivatives of the Stirling rate gives $e_2\equiv 0$, $e_3\equiv 0$ and $e_4\equiv -3b^8/(256\,a^2c^2\,x^2(b^4-x)^2)$ identically in (a,b,x). The identity therefore holds for every rational θ_0 treated in Theorem 7 of the main paper. At $\theta_0=\frac{1}{2}$, where the ridge is $y=2x$, the two steps can be displayed. (i) Differentiating the envelope identity once more gives $\partial_y^2 s=-\partial_y\log\eta=-(\Sigma^{-1})_{22}$ at the saddle, by the Legendre duality of Lemma S3(a), so that on the ridge $\partial_y^2 s=-16/(x(16-x))$ by Eq. (S1.17), while Eq. (S1.26) gives $\partial_y^2\gamma=\frac{1}{2x}+\frac{1}{2x}+\frac{1}{2(16-x)}+\frac{1}{2(16-x)}=16/(x(16-x))$ there. Hence $e_2=\frac{1}{2}\,\partial_y^2E=0$ on the ridge, and $e_3=0$ by the evenness of Lemma S3(b). (ii) Differentiating the Legendre relation twice more and evaluating on the ridge,

$$\partial_y^4 s\,\big|_{y=2x}=\frac{16\,(3x^2-48x-128)}{x^3(16-x)^3}\,,\qquad \partial_y^4\gamma\,\big|_{y=2x}=\frac{8\,(3x^2-48x+256)}{x^3(16-x)^3}\,,\tag{S1.32}$$

whose sum is $-72/(x^2(16-x)^2)$, so that $e_4=\frac{1}{24}\,\partial_y^4E=-3/(x^2(16-x)^2)=-q_4(x)$, as in Lemma 2 of the main paper, with the exact values $q_4(8)=\frac{3}{4096}$, $q_4(5/2)=\frac{16}{6075}$ and $q_4(32/3)=\frac{243}{262144}$.

The saddle series used in this computation exists and converges by Lemma S3(a), since the saddle is an analytic function of (x,y) on $\mathcal{U}$. The next lemma supplies the uniform remainder that Lemma 2 of the main paper, a statement about Taylor coefficients, does not contain.

Lemma S4 (Uniform quartic remainder)**.** *Let $K \subset (0,16)$ be compact, let ϵ_K be as in Lemma S3(b), and let $q_K = \min_K q_4 > 0$ with $q_4(x) = 3/(x^2(16-x)^2)$. There are $M_K < \infty$ and $\epsilon'_K \in (0, \epsilon_K]$ such that for all $x \in K$,*

$$\left|E(x, 2x+\epsilon) - E^* + q_4(x)\,\epsilon^4\right| \le M_K\,\epsilon^6 \quad (|\epsilon| \le \epsilon_K)\,, \qquad E(x, 2x+\epsilon) - E^* \le -\tfrac{1}{2}\,q_4(x)\,\epsilon^4 \quad (|\epsilon| \le \epsilon'_K)\,. \tag{S1.33}$$

Proof. By Lemma S3(b), $\epsilon \mapsto E(x, 2x+\epsilon)$ is analytic and even on $[-\epsilon_K, \epsilon_K]$ for each $x \in K$, and E is analytic on the compact tube $\mathcal{T}_K$, so $M_K = \sup_{\mathcal{T}_K} |\partial_\epsilon^6 E|/6!$ is finite. By Lemma 1 of the main paper $E(x, 2x) = E^*$, by Lemma 2 of the main paper the coefficients of ϵ^2 and ϵ^3 vanish and that of ϵ^4 is $-q_4(x)$, and by evenness the coefficients of ϵ and ϵ^5 vanish. Taylor's theorem with remainder of order six gives the first bound. With $\epsilon'_K = \min\left(\epsilon_K, (q_K/(2M_K))^{1/2}\right)$, $M_K\epsilon^6 \le \frac{1}{2}q_K\epsilon^4 \le \frac{1}{2}q_4(x)\epsilon^4$ for $|\epsilon| \le \epsilon'_K$, which gives the second. □

Proposition S2 (The transverse sum)**.** *Let $K \subset (0,16)$ be compact. There are M_K and n_K such that for all $n \ge n_K$ and all integers m with $x = m/n \in K$,*

$$\begin{aligned} \left| n^{7/4}\,\mathrm{row}_n(m) - p(x) \int_{\mathbb{R}} e^{-q_4(x)t^4} dt \right| &\le M_K\, n^{-1/4}\,, \\ \int_{\mathbb{R}} e^{-q_4(x)t^4} dt = \frac{2\,\Gamma(\frac{5}{4})}{q_4(x)^{1/4}} &= \frac{2\,\Gamma(\frac{5}{4})}{3^{1/4}} \sqrt{x(16-x)}\,. \end{aligned} \tag{S1.34}$$

Equivalently, $\mathrm{row}_n(m) = n^{-7/4} f(x)\left(1 + \mathcal{O}_K(n^{-1/4})\right)$ *uniformly for $x = m/n \in K$, with*

$$f(x) = p(x) \cdot 2\,\Gamma(\tfrac{5}{4})\,q_4(x)^{-1/4} = \frac{\sqrt{2\pi}\;2\,\Gamma(\frac{5}{4})}{64 \cdot 3^{1/4}}\;\frac{1}{\sqrt{x(16-x)}}\,. \tag{S1.35}$$

Proof. Fix m with $x = m/n \in K$ and write $\mathrm{row}_n(m) = \sum_{j \in \mathbb{Z}} S_n(m, 2m+j)$, where $S_n = 0$ off $\mathcal{L}_n$. Let ϵ'_K be as in Lemma S4 and put $t_j = jn^{-3/4}$, so that $\epsilon = j/n = t_j n^{-1/4}$ and $n\epsilon^4 = t_j^4$.

Near range $|j| \le \epsilon'_K n$. By Eq. (S1.30) and Eq. (S1.33), for $n \ge n_K$,

$$\frac{n^{5/2} S_n(m, 2m+j)}{p(x)} = e^{-q_4(x)t_j^4 + \varsigma}\,(1 + \tilde{\xi})\,, \qquad |\varsigma| \le M_K t_j^6 n^{-1/2}, \qquad |\tilde{\xi}| \le M_K\left(|t_j| n^{-1/4} + n^{-1/2}\right), \tag{S1.36}$$

where $\varsigma = n[E(x, 2x+\epsilon) - E^*] + q_4 t_j^4$ and $1 + \tilde{\xi} = (1 + \xi_n)A(x,\epsilon)/A(x,0)$. By the second part of Eq. (S1.33), $-q_4 t_j^4 + \varsigma \le -\frac{1}{2}q_4 t_j^4$, and also $-q_4 t_j^4 \le -\frac{1}{2}q_4 t_j^4$. Using $|e^\alpha - e^\beta| \le |\alpha - \beta| \max(e^\alpha, e^\beta)$ for real α, β,

$$\left| \frac{n^{5/2} S_n(m, 2m+j)}{p(x)} - e^{-q_4(x)t_j^4} \right| \le \left(|\varsigma| + |\tilde{\xi}|\right) e^{-q_4 t_j^4/2} \le M_K n^{-1/4}(1 + |t_j| + t_j^6) e^{-q_K t_j^4/2} \le M_K n^{-1/4} \Lambda(t_j), \tag{S1.37}$$

with $\Lambda(t) = M'_K e^{-q_K t^4/4}$ and $M'_K = \sup_t (1 + |t| + t^6) e^{-q_K t^4/4}$. The function Λ is even and decreasing on $[0, \infty)$. For such a function and any $h_0 > 0$, $h_0 \sum_{j \in \mathbb{Z}} \Lambda(jh_0) \le h_0 \Lambda(0) + \int_{\mathbb{R}} \Lambda$, and likewise $\left|h_0 \sum_{j \in \mathbb{Z}} e^{-q_4 (jh_0)^4} - \int_{\mathbb{R}} e^{-q_4 t^4} dt\right| \le h_0$. Taking $h_0 = n^{-3/4}$ and summing Eq. (S1.37) over $|j| \le \epsilon'_K n$,

$$\begin{aligned} \left| n^{7/4} \sum_{|j| \le \epsilon'_K n} \frac{S_n(m, 2m+j)}{p(x)} - \int_{\mathbb{R}} e^{-q_4 t^4} dt \right| &\le M_K n^{-1/4} \left(n^{-3/4}\Lambda(0) + \int \Lambda\right) + n^{-3/4} + n^{-3/4} \sum_{|j| > \epsilon'_K n} e^{-q_4 t_j^4} \\ &\le M_K n^{-1/4}, \end{aligned} \tag{S1.38}$$

where the last sum is at most $2\int_{\epsilon'_K n^{1/4}-1}^{\infty} e^{-q_K t^4}\,dt \le M_K e^{-n^{1/4}}$ for $n \ge n_K$, by monotonicity.

Far range $|j| > \epsilon'_K n$. By Eq. (S1.11), each such term satisfies $S_n(m, 2m{+}j) \le \exp\{-\epsilon'^4_K n/131072\}$, and there are at most $32n+1$ of them, so

$$n^{7/4} \sum_{|j|>\epsilon'_K n} S_n(m, 2m+j) \le 33\, n^{11/4} e^{-\epsilon'^4_K n/131072} \le M_K n^{-1/4}\,. \tag{S1.39}$$

Adding Eq. (S1.38) multiplied by $p(x) \le \max_K p$ and Eq. (S1.39) gives Eq. (S1.34). The value of the integral follows from the substitution $u = q_4 t^4$: $\int_0^\infty e^{-q_4 t^4} dt = \frac{1}{4} q_4^{-1/4}\Gamma(\frac{1}{4}) = q_4^{-1/4}\Gamma(\frac{5}{4})$, and $q_4^{-1/4} = (x^2(16-x)^2/3)^{1/4} = \sqrt{x(16-x)}/3^{1/4}$. □

At finite n the summand need not be largest on the ridge: the factor $1+\tilde{\xi}$ of Eq. (S1.36) allows $S_n(m, 2m+j)$ to exceed $S_n(m, 2m)$ by a relative amount of order $1/n$ at offsets $|j|$ of order $\sqrt{n}$, which is why the envelope constant in Eq. (S1.37) exceeds one. At $n = 32$, $m = 8n$, for example, exact evaluation gives $S_n(m, 2m+j) > S_n(m, 2m)$ for $1 \le |j| \le 31$ and not for $|j| = 32$, with the largest ratio, 1.0059, at $|j| = 23$.

The ends of the ridge. As $x \to 0$ or $x \to 16$ the tilt ζ^* tends to 0 or ∞, $\det \Sigma \to 0$, $q_4(x) \to \infty$ and $p(x) \to \infty$, so the constants of Lemmas S2–S4 and Proposition S2 degrade and no statement above is uniform on $(0, 16)$. The rows near the ends are bounded directly, without the local limit theorem and without Lemma 2 of the main paper, by an anti-concentration bound for the ζ-coefficients combined with the exact identity Eq. (S1.8).

Lemma S5 (Anti-concentration)**.** *Let X be a sum of finitely many independent Bernoulli variables and $V = \operatorname{Var} X$. Then $\sup_k \mathbb{P}(X = k) \le \frac{6}{5}(1+V)^{-1/2}$. Consequently, if $\mathcal{G}(\zeta) = \prod_v (1+\zeta\omega_v)^{nc_v}$ with all $\omega_v > 0$, then for every $\zeta_0 > 0$ and every integer k,*

$$[\zeta^k]\,\mathcal{G} \;\le\; \frac{6}{5}\,\frac{\mathcal{G}(\zeta_0)\,\zeta_0^{-k}}{\sqrt{1+V}}\,, \qquad V = n\sum_v c_v \pi_v(1-\pi_v)\,, \quad \pi_v = \frac{\zeta_0\omega_v}{1+\zeta_0\omega_v}\,. \tag{S1.40}$$

Proof. With π_r the success probabilities of the summands of X, $\mathbb{P}(X = k) = (2\pi)^{-1}\int_{-\pi}^{\pi}\prod_r(1-\pi_r+\pi_r e^{iu})e^{-iku}du$ and, as in Eq. (S1.20), $\prod_r |1-\pi_r+\pi_r e^{iu}| \le e^{-V(1-\cos u)} \le e^{-2Vu^2/\pi^2}$ on $[-\pi, \pi]$. Hence $\mathbb{P}(X = k) \le (2\pi)^{-1}\int_{\mathbb{R}} e^{-2Vu^2/\pi^2}du = \sqrt{\pi/(8V)}$. If $V \le \frac{11}{25}$ then $\frac{6}{5}(1+V)^{-1/2} \ge 1 \ge \mathbb{P}(X = k)$, and if $V \ge \frac{11}{25}$ then $\pi(1+V)/(8V) \le \frac{\pi}{8}(1+\frac{25}{11}) < \frac{36}{25}$, so $\sqrt{\pi/(8V)} < \frac{6}{5}(1+V)^{-1/2}$. For Eq. (S1.40), $[\zeta^k]\,\mathcal{G}\cdot\zeta_0^k/\mathcal{G}(\zeta_0)$ is the probability that a sum of $16n$ independent Bernoulli variables, nc_v of them with parameter π_v, equals k. □

Lemma S6 (Endpoint rows)**.** *For all $n \ge 1$ and $(m, \ell) \in \mathcal{L}_n$, $S_n(16n-m, 32n-\ell) = S_n(m, \ell)$, so $\mathrm{row}_n(16n-m) = \mathrm{row}_n(m)$. There is an absolute constant C_0 such that for all $n \ge 1$,*

$$\mathrm{row}_n(0) \le C_0\, n^{-3/2}\,, \qquad \mathrm{row}_n(m) \le C_0 \min\left(n^{-3/2},\, n^{-5/4}m^{-1/2}\right) \quad (1 \le m \le 8n)\,. \tag{S1.41}$$

Consequently, for every $\kappa \in (0, 8]$ and $n \ge 1$,

$$\sum_{0\le m\le \kappa n} \mathrm{row}_n(m) + \sum_{(16-\kappa)n\le m\le 16n} \mathrm{row}_n(m) \;\le\; 8\,C_0\left(\sqrt{\kappa}\, n^{-3/4} + n^{-1}\right). \tag{S1.42}$$

Proof. Mirror symmetry. The involution $x \mapsto U - x$ on allocations preserves $\prod_v \binom{U_v}{x_v}$ and maps (m, ℓ) to $(16n - m, 32n - \ell)$ by Eq. (S1.3), so $\phi_n(16n - m, 32n - \ell) = \phi_n(m, \ell)$. In Eq. (S1.4) the substitution exchanges the second and third brackets and swaps the arguments of the first, and D is symmetric. Hence S_n is invariant and it suffices to treat $m \le 8n$.

The row $m = 0$. Here $\ell = 0$ and $\mathrm{row}_n(0) = 2^{64n} D(0, 16n) D(0, 0) D(32n, 32n)$ with $D(0, 16n) = 1/(16n + 1)$ and $D(0, 0) = 1$. By Eq. (S1.25),

$$D(32n, 32n) \le e^{\varrho(32n,32n)}\, b_0(32n, 32n)\, e^{1/(192n)} = 2^{-64n} \frac{1}{16}\sqrt{\frac{2\pi}{n}}\, e^{1/(192n)}\,, \tag{S1.43}$$

so $\mathrm{row}_n(0) \le \sqrt{2\pi}\, e^{1/192} n^{-3/2}/256 \le n^{-3/2}$.

Row representation, $1 \le m \le 8n$. Put $\sigma^* = m/(16n) \in (0, \frac{1}{2}]$. Writing the second and third brackets of Eq. (S1.4) as Beta integrals over θ and ρ and summing over ℓ under the integral,

$$\mathrm{row}_n(m) = 2^{64n} D(m, 16n - m) \int_0^1\!\!\int_0^1 \theta^{4m} \rho^{32n-4m} (1 - \rho)^{32n}\, [\zeta^m]\, G(\zeta, \tilde\eta)^n\, d\theta\, d\rho\,, \qquad \tilde\eta = \frac{(1 - \theta)\rho}{\theta(1 - \rho)}\,, \tag{S1.44}$$

since $\sum_\ell \phi_n(m, \ell)\tilde\eta^{\,\ell} = [\zeta^m]G(\zeta, \tilde\eta)^n$. For $\theta, \rho \in (0, 1)$ apply Lemma S5 with $\omega_v = \tilde\eta^{\,v}$ and $\zeta_0 = \sigma^*\theta^4/((1 - \sigma^*)\rho^4)$. Then $\zeta_0\tilde\eta^{\,v} = \sigma^* f_v(\theta)/((1 - \sigma^*) f_v(\rho))$, so that, with $p_v = p_v(\sigma^*, \theta, \rho)$ as in Eq. (S1.6),

$$1 + \zeta_0\tilde\eta^{\,v} = \frac{p_v}{(1 - \sigma^*) f_v(\rho)}\,, \qquad \pi_v = \frac{\sigma^* f_v(\theta)}{p_v}\,, \qquad G(\zeta_0, \tilde\eta) = \frac{\prod_v p_v^{c_v}}{(1 - \sigma^*)^{16} \rho^{32} (1 - \rho)^{32}}\,, \tag{S1.45}$$

using $\prod_v f_v(\rho)^{c_v} = \rho^{32}(1-\rho)^{32}$. The monomials collapse as in the proof of Lemma S1: $\theta^{4m}\rho^{32n-4m}(1-\rho)^{32n} G(\zeta_0, \tilde\eta)^n \zeta_0^{-m} = \prod_v p_v^{nc_v}\, (\sigma^*)^{-m}(1 - \sigma^*)^{-(16n-m)}$. By Eq. (S1.25) with $(\alpha, \beta) = (m, 16n - m)$, $D(m, 16n - m)(\sigma^*)^{-m}(1 - \sigma^*)^{-(16n-m)} \le e^{1/6} b_0(m, 16n - m) \le e^{1/6}\sqrt{2\pi m}/(16n)$. With Eq. (S1.7) and Eq. (S1.40), therefore,

$$\mathrm{row}_n(m) \le \frac{\sqrt{m}}{4n} \int_0^1\!\!\int_0^1 \frac{e^{-16n\, D(Q\|P(\sigma^*,\theta,\rho))}}{\sqrt{1 + V(\theta, \rho)}}\, d\theta\, d\rho\,, \qquad V(\theta, \rho) = n\sigma^*(1 - \sigma^*) \sum_v c_v \frac{f_v(\theta) f_v(\rho)}{p_v^2}\,, \tag{S1.46}$$

using $\frac{6}{5} e^{1/6}\sqrt{2\pi}/16 < \frac{1}{4}$.

Variance floor. If $|\rho - \frac{1}{2}| \le \frac{1}{8}$ then $f_v(\rho) \ge (3/8)^4$ for every v, and $p_v \le 1$, so $V \ge n\sigma^*(1 - \sigma^*)(3/8)^4 \sum_v c_v f_v(\theta) = n\sigma^*(1 - \sigma^*)(3/8)^4 \ge c_1 m$ with $c_1 = 81/131072$, since $n\sigma^* = m/16$ and $1 - \sigma^* \ge \frac{1}{2}$.

Two-moment lower bound for the divergence. Let $\bar\theta = \frac{1}{2} - \theta$, $\bar\rho = \frac{1}{2} - \rho$, $\Delta = \sigma^*\bar\theta + (1 - \sigma^*)\bar\rho$ and $\Xi = \sigma^*\bar\theta^2 + (1 - \sigma^*)\bar\rho^2 \ge 0$. Since $P(\sigma^*, \theta, \rho)$ is the mixture of $\mathrm{Bin}(4, \frac{1}{2} + \bar\theta)$ and $\mathrm{Bin}(4, \frac{1}{2} + \bar\rho)$ with weights $\sigma^*, 1 - \sigma^*$, its first two factorial moments are $\sum_v vP_v = 4(\frac{1}{2} + \Delta)$ and $\sum_v v(v - 1)P_v = 12(\frac{1}{4} + \Delta + \Xi)$, against 2 and 3 for Q. As $0 \le v \le 4$ and $0 \le v(v - 1) \le 12$, $\|P - Q\|_1 \ge |\Delta|$ and $\|P - Q\|_1 \ge |\Delta + \Xi| \ge \Xi - |\Delta|$, and Pinsker's inequality gives $D(Q\|P) \ge \frac{1}{2}\max\left(\Delta^2, (\Xi - |\Delta|)_+^2\right)$. If $|\Delta| \ge \Xi/2$ this is at least $\frac{1}{2}\Delta^2 \ge \frac{1}{4}\Delta^2 + \frac{1}{16}\Xi^2$, and if $|\Delta| < \Xi/2$ it is at least $\frac{1}{8}\Xi^2 \ge \frac{1}{16}\Xi^2 + \frac{1}{4}\Delta^2$. In both cases

$$D(Q \,\|\, P(\sigma^*, \theta, \rho)) \;\ge\; \tfrac{1}{4}\Delta^2 + \tfrac{1}{16}\Xi^2\,. \tag{S1.47}$$

The row bound. With $P = P(\sigma^*, \theta, \rho)$, split the square in Eq. (S1.46) at $|\bar\rho| = \frac{1}{8}$. Where $|\bar\rho| > \frac{1}{8}$, $\Xi \ge (1 - \sigma^*)\bar\rho^2 \ge 2^{-7}$, so $16n\, D(Q\|P) \ge n\,\Xi^2 \ge 2^{-14} n$ by Eq. (S1.47), and with $(1 + V)^{-1/2} \le 1$ this part of Eq. (S1.46) is at most $\frac{\sqrt{m}}{4n} e^{-n/16384} \le n^{-1/2} e^{-n/16384}$ for $m \le 8n$. Where $|\bar\rho| \le \frac{1}{8}$, use

$(1+V)^{-1/2} \le (c_1 m)^{-1/2}$, $16n\, D(Q\|P) \ge 4n\Delta^2 + n\Xi^2 \ge 4n\Delta^2 + n\sigma^{*2}\bar\theta^4$, and the change of variables $(\bar\theta, \bar\rho) \mapsto (\bar\theta, \Delta)$ with $d\Delta = (1-\sigma^*)d\bar\rho \ge \frac{1}{2}d\bar\rho$:

$$\begin{aligned}\iint_{|\bar\rho|\le 1/8} \frac{e^{-16n\, D(Q\|P)}}{\sqrt{1+V}}\, d\theta\, d\rho &\le \frac{2}{\sqrt{c_1 m}} \int_{\mathbb{R}} e^{-4n\Delta^2} d\Delta \int_{-1/2}^{1/2} e^{-n\sigma^{*2}\bar\theta^4} d\bar\theta \\ &\le \frac{2}{\sqrt{c_1 m}} \sqrt{\frac{\pi}{4n}}\, \min\Big(1,\ 2\Gamma(\tfrac{5}{4})\big(n\sigma^{*2}\big)^{-1/4}\Big).\end{aligned} \tag{S1.48}$$

Since $(n\sigma^{*2})^{-1/4} = (256n/m^2)^{1/4} = 4n^{1/4}m^{-1/2}$, the central part of Eq. (S1.46) is at most $C' \min(n^{-3/2}, n^{-5/4}m^{-1/2})$ with an absolute constant C'. The outer part obeys the same bound for $m \le 8n$, since $n^{-1/2}e^{-n/16384} \le C' n^{-5/4}(8n)^{-1/2}$ as soon as $C' \ge \sqrt{8}\sup_n n^{5/4}e^{-n/16384}$, which is finite. This proves Eq. (S1.41).

Strips. By Eq. (S1.41), $\sum_{0\le m\le \kappa n} \mathrm{row}_n(m) \le C_0 n^{-3/2}(1+\sqrt{n}) + C_0 n^{-5/4} \sum_{\sqrt{n}<m\le \kappa n} m^{-1/2} \le 2C_0 n^{-1} + 2C_0\sqrt{\kappa}\, n^{-3/4}$, using $\sum_{m\le M} m^{-1/2} \le 2\sqrt{M}$, and the mirror strip contributes the same. □

In Eq. (S1.41) the bound $n^{-3/2}$ is the full prefactor of the three brackets, attained for m of order one, where the row is a fixed window of the lattice. For m large the bound $n^{-5/4}m^{-1/2} = n^{-7/4}x^{-1/2}$ is the order of $f(x)$ in Eq. (S1.35) as $x \to 0$, since $f(x)\sqrt{x}$ is bounded there, and the strip bound $\sqrt{\kappa}\, n^{-3/4}$ matches the tail $\int_0^\kappa f = \mathcal{O}(\sqrt{\kappa})$ of the ridge integral.

Proof of Theorem 6 of the main paper. Fix $\kappa \in (0,1]$ and let $K = [\kappa, 16-\kappa]$. By Eq. (S1.5),

$$Z(n)\, 2^{64n} = \sum_{\kappa n \le m \le (16-\kappa)n} \mathrm{row}_n(m) \ +\ R_n(\kappa)\,, \qquad 0 \le R_n(\kappa) \le 8\, C_0\big(\sqrt{\kappa}\, n^{-3/4} + n^{-1}\big)\,, \tag{S1.49}$$

where $R_n(\kappa)$ is the sum of the rows with $m < \kappa n$ or $m > (16-\kappa)n$, bounded by Eq. (S1.42). By Proposition S2, for $n \ge n_K$ the first sum equals $n^{-3/4} \cdot \frac{1}{n}\sum_{m/n\in K} f(m/n)\,\big(1 + \mathcal{O}_K(n^{-1/4})\big)$. Since f is continuously differentiable on K, $\big|\frac{1}{n}\sum_{m:\, m/n\in K} f(m/n) - \int_K f\big| \le (16\max_K |f'| + 2\max_K f)/n$. Hence

$$\limsup_{n\to\infty} \Big| n^{3/4} Z(n)\, 2^{64n} - \int_0^{16} f(x)\, dx \Big| \ \le\ 8\, C_0\sqrt{\kappa} + \int_{[0,\kappa]\cup[16-\kappa,16]} f(x)\, dx\,, \tag{S1.50}$$

and the right side tends to zero as $\kappa \to 0$ because f is integrable on $[0,16]$ by Eq. (S1.35). The left side does not depend on κ, so it is zero. With $\int_0^{16} dx/\sqrt{x(16-x)} = \pi$ (substitute $x = 8(1-\cos\varphi)$),

$$\int_0^{16} f(x)\, dx = \frac{\sqrt{2\pi}\; 2\,\Gamma(\frac{5}{4})}{64 \cdot 3^{1/4}} \cdot \pi = \frac{\sqrt{2}\, \pi^{3/2}\, \Gamma(\frac{5}{4})}{32 \cdot 3^{1/4}} = C\,, \tag{S1.51}$$

which is Theorem 6 of the main paper: $Z(n) = C\, n^{-3/4}\, 2^{-64n}(1 + o(1))$. □

The three factors described informally before Theorem 6 in the main paper appear in Eq. (S1.51) as π (the ridge integral), $2\Gamma(\frac{5}{4})q_4^{-1/4}$ (the transverse sum, Proposition S2) and $p(x)$ (the local limit theorem and the Stirling prefactors, Lemma S3(c)). The proof gives the rate $\mathcal{O}_K(n^{-1/4})$ on compact ridge segments and $\mathcal{O}(\sqrt{\kappa}) + \mathcal{O}(n^{-1/4})$ for the strips, but it does not determine the order of the total correction, which Table 2 of the main paper indicates is $n^{-1/4}$.

Remark S1 (The stratum family)**.** *For Theorem 7 of the main paper, with $\theta_0 = a/b$, $c = b - a$ and data $U_v = n\, d_v$, $d_v = \binom{4}{v} a^{4-v} c^v$, $n = N/b^4$, the same proof applies with the following changes. The sums Eq.* (S1.3) *become $\sum_v d_v = b^4$, $\sum_v (4-v) d_v = 4ab^3$, $\sum_v v d_v = 4cb^3$, the reference law is $Q = \mathrm{Bin}(4, 1-\theta_0)$, the ridge is $y = (4c/b)x$ with tilt $(\zeta^*, \eta^*) = (x/(b^4 - x), 1)$, and Eq.* (S1.7) *reads $\prod_v \bar{q}_v^{-nd_v} \prod_v p_v^{nd_v} = e^{-b^4 n D(Q\|P)}$ with $\bar{q}_v = f_v(\theta_0)$, so that E^* is replaced by the rate $4N[\theta_0 \log \theta_0 + (1-\theta_0)\log(1-\theta_0)]$ of the theorem. Identity Eq.* (S1.8)*, Lemma S1 and its moment identity (now centered at $1-\theta_0$), the saddle chart, the Stirling prefactors and Lemma S5 go through with constants depending on (a, b). So does the local limit theorem: the step lattice is still $\mathbb{Z}^2$ since $d_0, d_1 \geq 1$, and M becomes $\sum_v d_v \binom{1}{v}\binom{1}{v}^\top$, positive definite by the Cauchy–Schwarz inequality. Two steps change form. First, for $a \neq c$ the exponent is no longer even in ϵ, so Lemma S4 holds with remainder $M_K|\epsilon|^5$ from the vanishing of the quadratic and cubic coefficients in the q_4 identity above. In Eq.* (S1.36) *this gives $|\varsigma| \leq M_K |t_j|^5 n^{-1/4}$, which leaves Proposition S2 and its rate unchanged. Second, the mirror symmetry of Lemma S6 is lost, and the strip near $m = b^4 n$ is treated instead by exchanging the roles of the two components, $(\sigma, \theta, \rho) \mapsto (1-\sigma, \rho, \theta)$, which exchanges the second and third brackets of the weight. The variance floor and the two-moment bound then have θ_0-dependent centers and constants. The prefactor assembly changes as follows. The matrix $\mathsf{M} = \sum_v d_v \binom{1}{v}\binom{1}{v}^\top$ above has determinant $4acb^6$, and at the ridge tilt every $\pi_v = \varpi = x/b^4$, so that Eq.* (S1.17) *becomes $\sqrt{\det \Sigma} = 2\sqrt{ac}\, x(b^4 - x)/b^5$. On the ridge the three pairs in Eq.* (S1.26) *become $(x, b^4 - x)$, $(4ax/b,\, 4cx/b)$ and $(4a(b^4 - x)/b,\, 4c(b^4 - x)/b)$, which gives $W = \sqrt{2\pi}\, \pi ac/(2b^8)$, again independent of x, and therefore*

$$p(x; \theta_0) = \frac{W}{2\pi\sqrt{\det \Sigma}} = \frac{\sqrt{2\pi ac}}{8\, b^3\, x(b^4 - x)}\,, \qquad q_4(x; \theta_0) = \frac{3b^8}{256\, a^2 c^2\, x^2 (b^4 - x)^2} \tag{S1.52}$$

in place of Eq. (S1.29) *and $q_4(x)$. Since $n^{-3/4} = b^3 N^{-3/4}$ and $\int_0^{b^4} dx/\sqrt{x(b^4 - x)} = \pi$, the analog of Eq.* (S1.51) *is*

$$C_N(\theta_0) = b^3 \int_0^{b^4} p(x; \theta_0)\, \frac{2\,\Gamma(\frac{5}{4})}{q_4(x; \theta_0)^{1/4}}\, dx = \frac{\sqrt{2}\, \pi^{3/2}\, \Gamma(\frac{5}{4})}{3^{1/4}} \cdot \frac{ac}{b^2} = \frac{\sqrt{2}\; 3^{3/4}\, \pi^{3/2}\, \Gamma(\frac{1}{4})}{12}\, \theta_0(1 - \theta_0)\,, \tag{S1.53}$$

which is the constant of Theorem 7 of the main paper.

S1.3 Localization for the Gaussian location mixture

We prove Proposition 1 of the main paper (Section 5.3), the population-ray asymptotics of the evidence of the two-component Gaussian location mixture, with a bound $\mathcal{O}(n^{-1/4})$ on the relative correction. The geometry is that of the ridge of the binomial mixture (Section 5.1 of the main paper). The Kullback–Leibler divergence K vanishes on a fiber and is quadratic in one transverse direction and quartic in the other, so the leading constant is a fiber integral of two transverse Laplace factors. The neighborhoods of the boundary strata $\alpha \in \{0, 1\}$ contribute $\mathcal{O}(n^{-1})$, and the region away from the zero set of K an exponentially small amount, against $n^{-3/4}$ from the fiber. The ends of the fiber, where the quartic coefficient $\alpha^2(1-\alpha)^2/4$ vanishes, enter the leading term only through the integrable density $d\alpha/\sqrt{\alpha(1-\alpha)}$.

Setting. Fix $s_0 > 0$ and $\mu_0 \in \mathbb{R}$. The truth is $q(y) = N(y; \mu_0, 1)$, and $\varphi(x) = e^{-x^2/(2s_0^2)}/(s_0\sqrt{2\pi})$ denotes the $N(0, s_0^2)$ density, with $\varphi_0 = \varphi(\mu_0)$ and $\|\varphi\|_\infty = 1/(s_0\sqrt{2\pi})$. Model 1 is $N(y; \mu, 1)$ with

$\mu \sim N(0, s_0^2)$. Model 2 is $p_\theta(y) = \alpha\, N(y; \mu_1, 1) + (1-\alpha)\, N(y; \mu_2, 1)$, $\theta = (\alpha, \mu_1, \mu_2)$, with α uniform on $[0,1]$ and μ_1, μ_2 iid $N(0, s_0^2)$, independent of α. For $j \in \{1,2\}$ let π_j denote the prior of model j. For real $n > 0$ the population-ray evidence is $Z_j(n) = \int e^{nL_j(\theta)}\, d\pi_j(\theta)$ with $L_j(\theta) = \int q \log p_\theta$, as in Section 5.3 of the main paper; for model 1 the parameter is $\theta = \mu$ and $p_\theta = N(\cdot\,; \mu, 1)$. Let $L_0 = \int q \log q = -\frac{1}{2}\log(2\pi) - \frac{1}{2}$. Since $L_0 - L_j(\theta)$ is a Kullback–Leibler divergence, $L_0 = \sup_\theta L_j(\theta)$ for both models, and for model 2 we write

$$K(\theta) \;=\; \int q\, \log\frac{q}{p_\theta} \;=\; L_0 - L_2(\theta) \;\geq\; 0\,, \qquad e^{-nL_0} Z_2(n) \;=\; \int e^{-nK(\theta)}\, d\pi_2(\theta)\,. \tag{S1.54}$$

Throughout, $n \to \infty$ through real values and $\mathbb{E}$ denotes expectation over $z \sim N(0,1)$.

Proposition S3. *For every $s_0 > 0$ and $\mu_0 \in \mathbb{R}$:*

(i) $e^{-nL_0} Z_1(n) = (1+ns_0^2)^{-1/2} \exp\{-n\mu_0^2/(2(1+ns_0^2))\} = \varphi_0\sqrt{2\pi/n}\,(1+\mathcal{O}(1/n))$*;*

(ii) $n^{3/4}\, e^{-nL_0} Z_2(n) = C_2 + \mathcal{O}(n^{-1/4})$ *with* $C_2 = \pi^{3/2}\,\Gamma(\frac{1}{4})\,\varphi_0^2 = \sqrt{\pi}\,\Gamma(\frac{1}{4})\, e^{-\mu_0^2/s_0^2}/(2s_0^2)$*;*

(iii) $\mathrm{BF}(n) = Z_1(n)/Z_2(n) = c(s_0, \mu_0)\, n^{1/4}(1 + \mathcal{O}(n^{-1/4}))$ *with*

$$c(s_0, \mu_0) \;=\; \frac{\varphi_0\sqrt{2\pi}}{C_2} \;=\; \frac{2\, s_0\, e^{\mu_0^2/(2s_0^2)}}{\sqrt{\pi}\;\Gamma(1/4)}\,. \tag{S1.55}$$

The implied constants depend on (s_0, μ_0) only.

Part (ii) says that $e^{-nL_0} Z_2(n)$ is of exact order $n^{-3/4}$, with no logarithm, so that on this ray the learning coefficient is $\lambda = 3/4$ with multiplicity $m = 1$, the values stated in Section 5.3 of the main paper, in agreement with the real log canonical threshold of this model (Aoyagi, 2010)(Drton and Plummer, 2017, Example 2.1). Parts (ii) and (iii) contain Proposition 1 of the main paper. As in Section 5.1 of the main paper, the factor $C_2/\varphi_0^2 = \sqrt{2\pi} \cdot \Gamma(\frac{1}{4})/\sqrt{2} \cdot \pi$ is a product of three terms. The first two are the Gaussian transverse integral $\sqrt{2\pi}$ and the quartic transverse integral $\int e^{-v^4/4} dv = \Gamma(\frac{1}{4})/\sqrt{2}$, and the third is the mass $\int_0^1 d\alpha/\sqrt{\alpha(1-\alpha)} = \pi$ of the density that the α-dependent quartic width induces on the fiber.

Coordinates. Put $t_h = \mu_h - \mu_0$, so that $\theta \leftrightarrow (\alpha, t_1, t_2) \in [0,1] \times \mathbb{R}^2$ and $d\pi_2 = \varphi(\mu_0 + t_1)\varphi(\mu_0 + t_2)\, d\alpha\, dt_1\, dt_2$. The fiber-adapted coordinates of the main paper are

$$w = \alpha t_1 + (1-\alpha) t_2\,, \qquad s = t_1 - t_2\,, \qquad t_1 = w + (1-\alpha)s\,, \quad t_2 = w - \alpha s\,, \tag{S1.56}$$

with unit Jacobian for every α, and we write

$$u = \alpha(1-\alpha)s^2, \qquad \rho(\alpha) = \sqrt{\alpha(1-\alpha)}\,, \qquad m_k = \alpha t_1^k + (1-\alpha) t_2^k, \qquad |t| = \max(|t_1|, |t_2|)\,, \tag{S1.57}$$

and $Q_0 = \frac{w^2}{2} + \frac{u^2}{4}$. Thus $m_0 = 1$, $m_1 = w$, $m_2 = w^2 + u$, $|m_k| \leq |t|^k$, $|w| \leq |t|$ and $|s| \leq 2|t|$. The form Q_0 collects the two leading terms of the normal form of Section 5.3 of the main paper. The mean and variance of p_θ are

$$\int y\, p_\theta(y)\, dy = \mu_0 + w\,, \qquad \int (y - \mu_0 - w)^2 p_\theta(y)\, dy = 1 + u\,. \tag{S1.58}$$

With $z = y - \mu_0$ and $E_\tau(z) = e^{\tau z - \tau^2/2} = N(y; \mu_0 + \tau, 1)/q(y)$, the likelihood ratio is $p_\theta/q = 1 + A$ with $A = \alpha\,(E_{t_1} - 1) + (1-\alpha)(E_{t_2} - 1)$, so that $\mathbb{E}[A] = 0$, $1 + A > 0$, and

$$K = \mathbb{E}\big[-\log(1+A)\big] = \mathbb{E}\big[\psi(A)\big]\,, \qquad \psi(x) = x - \log(1+x) \ge 0 \quad (x > -1)\,. \tag{S1.59}$$

Since $\prod_i E_{\tau_i}(z) = \exp\{(\sum_i \tau_i)z - \sum_i \tau_i^2/2\}$ and $\mathbb{E}\,e^{cz} = e^{c^2/2}$,

$$\mathbb{E}\Big[\prod_{i=1}^{r} E_{\tau_i}\Big] = \exp\Big(\sum_{i<j} \tau_i\tau_j\Big)\,, \qquad \mathbb{E}\Big[\prod_{i=1}^{r}(E_{\tau_i} - 1)\Big] = \sum_{S \subseteq \{1,\dots,r\}} (-1)^{r-|S|} \exp\Big(\sum_{\substack{i<j\\ i,j\in S}} \tau_i\tau_j\Big). \tag{S1.60}$$

Zero set and regions. Let $\mathcal{Z} = \mathcal{R} \cup B_0 \cup B_1$ with $\mathcal{R} = \{t_1 = t_2 = 0\}$, $B_0 = \{\alpha = 0,\ t_2 = 0\}$, $B_1 = \{\alpha = 1,\ t_1 = 0\}$, subsets of $[0,1] \times \mathbb{R}^2$: the fiber $\mathcal{R}$ and the two boundary strata, which meet it at its ends. For $\delta > 0$ let

$$\mathrm{I}(\delta) = \{|t_1| \le \delta,\ |t_2| \le \delta\}, \quad \mathrm{II}(\delta) = \{\alpha \le \tfrac{1}{2},\ |t_1| > \delta \ge |t_2|\}, \quad \mathrm{II}'(\delta) = \{\alpha > \tfrac{1}{2},\ |t_2| > \delta \ge |t_1|\}, \tag{S1.61}$$

and let $\mathrm{III}(\delta)$ be the complement of $\mathrm{I} \cup \mathrm{II} \cup \mathrm{II}'$ in $[0,1] \times \mathbb{R}^2$, the union of $\{|t_1| > \delta,\ |t_2| > \delta\}$, $\{\alpha > \frac{1}{2},\ |t_1| > \delta,\ |t_2| \le \delta\}$ and $\{\alpha \le \frac{1}{2},\ |t_2| > \delta,\ |t_1| \le \delta\}$. In II the lighter component is displaced and the heavier one is within δ of the truth; the slabs in which the heavier component is displaced belong to III. The involution $(\alpha, t_1, t_2) \mapsto (1-\alpha, t_2, t_1)$ preserves K, the prior density, and I, and maps II' into $\{\alpha < \frac{1}{2},\ |t_1| > \delta,\ |t_2| \le \delta\} \subseteq \mathrm{II}$.

Lemma S7. *Let $\psi(x) = x - \log(1+x)$ and $\psi_4(x) = \psi(x) - \frac{x^2}{2} + \frac{x^3}{3}$ on $(-1, \infty)$, and $x_+ = \max(x, 0)$. Then $\psi_4'(x) = x^3/(1+x)$, and for all $x > -1$,*

$$0 \;\le\; \psi_4(x) \;\le\; \frac{x^4}{4}\max\Big(1, \frac{1}{1+x}\Big), \qquad \psi(x) \;\ge\; \frac{x^2}{2(1+x_+)}\,. \tag{S1.62}$$

Proof. The derivative $\psi_4'(x) = 1 - \frac{1}{1+x} - x + x^2 = \frac{x^3}{1+x}$ has the sign of x and $\psi_4(0) = 0$, so $\psi_4 \ge 0$ and $\psi_4(x) = \int_0^x r^3(1+r)^{-1}dr$. For $x \ge 0$ the integrand is at most r^3; for $-1 < x < 0$, $\psi_4(x) = \int_x^0 |r|^3(1+r)^{-1}dr \le (1+x)^{-1}x^4/4$. For the last inequality, if $x \ge 0$ then $\frac{d}{dx}\big[\psi(x) - \frac{x^2}{2(1+x)}\big] = \frac{x^2}{2(1+x)^2} \ge 0$ and the bracket vanishes at 0; if $-1 < x \le 0$ then $\psi(x) = \sum_{k\ge 2} |x|^k/k \ge x^2/2$. $\square$

Lemma S8. *(a) With $\alpha_1 = \alpha$, $\alpha_2 = 1-\alpha$, for all $(\alpha, t_1, t_2) \in [0,1] \times \mathbb{R}^2$,*

$$\mathbb{E}[A^2] = \sum_{a,b=1}^{2} \alpha_a\alpha_b(e^{t_a t_b} - 1) = \sum_{k\ge 1} \frac{m_k^2}{k!}\,, \qquad \mathbb{E}[A^3] = \sideset{}{'}\sum_{i,j,l\ge 0} \frac{m_{i+l}\,m_{i+j}\,m_{j+l}}{i!\,j!\,l!}\,, \tag{S1.63}$$

where $\sum'$ runs over the triples with at least two of i, j, l nonzero, so that all three indices $i+l$, $i+j$, $j+l$ are at least 1. Both series converge absolutely, and $\mathbb{E}|A|^r < \infty$ for every r.

(b) K is finite and continuous on $[0,1] \times \mathbb{R}^2$; $K(\alpha, -t_1, -t_2) = K(\alpha, t_1, t_2) = K(1-\alpha, t_2, t_1)$; and $K = t_2^2/2$ at $\alpha = 0$, $K = t_1^2/2$ at $\alpha = 1$, $K = t_1^2/2$ on $t_1 = t_2$.

(c) $K(\theta) = 0$ if and only if $\theta \in \mathcal{Z}$.

(d) For all (α, t_1, t_2), $K \ge -2\log b$ with $b = \sqrt{\alpha}\,e^{-t_1^2/8} + \sqrt{1-\alpha}\,e^{-t_2^2/8} \le \big(e^{-t_1^2/4} + e^{-t_2^2/4}\big)^{1/2}$.

Proof. (a) The closed form for $\mathbb{E}[A^2]$ is Eq. (S1.60) with $r = 2$ averaged over the component labels, and expanding $e^{t_a t_b} - 1$ and using $\sum_{a,b} \alpha_a \alpha_b (t_a t_b)^k = m_k^2$ gives the series. For $r = 3$, Eq. (S1.60) gives $\mathbb{E} \prod_c (E_{t_{ac}} - 1) = e^{x+y+\zeta} - e^x - e^y - e^\zeta + 2$ with $x = t_a t_b$, $y = t_b t_c$, $\zeta = t_a t_c$. Expanding, the subtracted exponentials remove exactly the terms $x^i y^j \zeta^l/(i!j!l!)$ with at most one nonzero index, and the label average of $x^i y^j \zeta^l = t_a^{i+l} t_b^{i+j} t_c^{j+l}$ is $m_{i+l} m_{i+j} m_{j+l}$. Absolute convergence follows from $|m_d| \le |t|^d$, and the moments of A are finite because $E_\tau(z)$ is lognormal.

(b) If $|t| \le T$ then $(2\pi)^{-1/2} e^{-(|z|+T)^2/2} \le p_\theta(\mu_0 + z) \le (2\pi)^{-1/2}$, so $|\log p_\theta(\mu_0 + z)| \le \frac{1}{2}\log(2\pi) + \frac{1}{2}(|z| + T)^2$, an integrable majorant uniform in θ. Since $p_\theta(y)$ is continuous in θ, dominated convergence gives finiteness and continuity of $K = L_0 - \int q \log p_\theta$. The symmetries follow from $E_{-\tau}(z) = E_\tau(-z)$ and from relabeling the components, and the boundary values are divergences between unit-variance normals.

(c) On $\mathcal{Z}$, $p_\theta = q$. Conversely, $K(\theta) = 0$ forces $p_\theta = q$, hence equal means and variances, and by Eq. (S1.58) $w = 0$ and $\alpha(1-\alpha)s^2 = 0$. If $s = 0$ then $t_1 = t_2 = w = 0$; if $\alpha = 0$ then $t_2 = w = 0$; if $\alpha = 1$ then $t_1 = w = 0$.

(d) By Jensen's inequality $-K = 2\,\mathbb{E}_q \log\sqrt{p_\theta/q} \le 2\log\int\sqrt{q\,p_\theta}$. By $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ the Bhattacharyya coefficient of q and the mixture is at most $\sqrt{\alpha}\int\sqrt{qN(\cdot;\mu_1,1)} + \sqrt{1-\alpha}\int\sqrt{qN(\cdot;\mu_2,1)}$, and $\int\sqrt{N(\cdot;\mu_0,1)N(\cdot;\mu,1)} = e^{-(\mu-\mu_0)^2/8}$. The last bound is the Cauchy–Schwarz inequality. $\square$

The normal form near the fiber. Let $\delta_0 = \frac{1}{3}$. On $\mathrm{I}(\delta_0)$,

$$|w| \le |t| \le \tfrac{1}{3}\,, \qquad |s| \le 2|t| \le \tfrac{2}{3}\,, \qquad |w| + |s| \le 3|t| \le 1\,, \qquad u \le s^2/4\,, \tag{S1.64}$$

and in the coordinates (α, w, s) the region $\mathrm{I}(\delta_0)$ lies in the compact set $\mathcal{C} = [0,1] \times \{|w| \le \frac{1}{3}\} \times \{|s| \le \frac{2}{3}\}$. By Eq. (S1.56), $\partial t_h/\partial w = 1$, hence $\partial_w m_d = d\, m_{d-1}$ for $d \ge 1$, and on the slice $w = 0$, where $t_1 = (1-\alpha)s$ and $t_2 = -\alpha s$,

$$m_d\big|_{w=0} = \alpha(1-\alpha)\, s^d\, \epsilon_d(\alpha)\,, \qquad \epsilon_d(\alpha) = (1-\alpha)^{d-1} - (-\alpha)^{d-1}\,, \qquad \epsilon_1 = 0, \quad \epsilon_2 = 1\,, \tag{S1.65}$$

and $|\epsilon_d| \le (1-\alpha)^{d-1} + \alpha^{d-1} \le 1$ for $d \ge 2$. A direct Taylor expansion of K leaves a remainder $\mathcal{O}(|t|^6)$, which does not control the sign of $K - Q_0$ near the ends of the fiber. When α is small compared with $|s|$, a term of order $|s|^6$ exceeds $u^2/4 = \alpha^2(1-\alpha)^2 s^4/4$. We therefore bound K from below by the minorant F of the next lemma, whose remainder terms each carry one of the factors w^2, $|w|u^2|s|$, $u^2 s^2$ and are thus controlled by Q_0 uniformly in $\alpha \in [0,1]$.

Lemma S9. *Let $F = \frac{1}{2}\mathbb{E}[A^2] - \frac{1}{3}\mathbb{E}[A^3]$, an entire function of (α, w, s) by Eq.* (S1.63) *and Eq.* (S1.60)*. Then:*

(a) $K \ge F$ on $[0,1] \times \mathbb{R}^2$.

(b) The Taylor expansion of F in (w, s) at the origin, whose coefficients are polynomials in α, has only terms of even total degree, quadratic part $w^2/2$, and quartic part $\frac{u^2}{4} - \frac{uw^2}{2} - \frac{3}{4}w^4$.

(c) On $\mathcal{C}$,

$$\begin{gathered} F = w^2\, G(\alpha, w, s) + w\, u^2 s\, P(\alpha, s) + \frac{u^2}{4} + u^2 s^2 H(\alpha, s)\,, \\ |H| \le C_H\,, \qquad |P| \le C_P\,, \qquad |G - \tfrac{1}{2}| \le C_G\,(|w| + |s|)^2\,, \end{gathered} \tag{S1.66}$$

with $C_H = \frac{e-5/2}{2} + \frac{e^3}{12}$, $C_P = (e-2) + 2e^3$, and $C_G = \frac{1}{2}\max\left(\|\partial_w^4 F\|_{\mathcal{C}}, \|\partial_w^3 \partial_s F\|_{\mathcal{C}}, \|\partial_w^2 \partial_s^2 F\|_{\mathcal{C}}\right) < \infty$, where $\|\cdot\|_{\mathcal{C}}$ denotes the supremum over $\mathcal{C}$ and the derivatives are taken in (w, s) at fixed α.

Proof. (a) Apply $\psi_4 \geq 0$ of Lemma S7 at $x = A(z)$ and take expectations, using Eq. (S1.59), $\mathbb{E}[A] = 0$ and the finiteness of the moments: $K - F = \mathbb{E}[\psi_4(A)] \geq 0$.

(b) In Eq. (S1.63) the total degree in t of a product $m_{d_1} \cdots m_{d_r}$ is $\sum d_i$, equal to $2k$ or $2(i+j+l)$, so only even degrees occur. Degree two comes from $k = 1$ only: $m_1^2/2 = w^2/2$. In degree four, $\frac{1}{2}\mathbb{E}[A^2]$ contributes $m_2^2/(2 \cdot 2!) = (w^2 + u)^2/4$ and $-\frac{1}{3}\mathbb{E}[A^3]$ contributes the three permutations of $(i, j, l) = (1, 1, 0)$, namely $-\frac{1}{3} \cdot 3\, m_1^2 m_2 = -w^2(w^2 + u)$. The sum is $\frac{u^2}{4} - \frac{uw^2}{2} - \frac{3}{4}w^4$.

(c) At fixed (α, s), Taylor's formula in w gives $F = F|_{w=0} + w\,(\partial_w F)|_{w=0} + w^2 G$ with $G = \int_0^1 (1-\tau)\, \partial_w^2 F(\alpha, \tau w, s)\, d\tau$, and the series Eq. (S1.63) may be evaluated and differentiated termwise, since $|m_d| \leq |t|^d$ and $\partial_w m_d = d\, m_{d-1}$ make both series and their termwise w-derivatives converge locally uniformly.

The w-free part. At $w = 0$, $m_1 = 0$, so in Eq. (S1.63) only products all of whose indices are at least 2 survive, and by Eq. (S1.65) each factor is $\alpha(1 - \alpha)s^d \epsilon_d$. Thus $\frac{1}{2}\mathbb{E}[A^2]|_{w=0} = \frac{1}{2}\sum_{k \geq 2} \alpha^2(1 - \alpha)^2 s^{2k} \epsilon_k^2/k! = \frac{u^2}{4} + u^2 s^2 \sum_{k \geq 3} s^{2k-6}\epsilon_k^2/(2\,k!)$. Each surviving term of $\mathbb{E}[A^3]|_{w=0}$ has three such factors, all of index at least 2, so that $i + j + l \geq 3$, and equals $\alpha(1 - \alpha) \cdot u^2 s^2 \cdot s^{2(i+j+l)-6} \epsilon_{i+l}\epsilon_{i+j}\epsilon_{j+l}/(i!j!l!)$. Hence $F|_{w=0} = \frac{u^2}{4} + u^2 s^2 H$ with $|H| \leq \frac{1}{2}\sum_{k \geq 3} \frac{1}{k!} + \frac{1}{12}\sum_{i,j,l \geq 0} \frac{1}{i!j!l!} = C_H$ for $|s| \leq 1$, using $|\epsilon_d| \leq 1$ and $\alpha(1 - \alpha) \leq \frac{1}{4}$.

The single-w part. Differentiating a product $\prod m_{d_i}$ in w replaces one factor m_d by $d\, m_{d-1}$, and at $w = 0$ a term survives only if every index present is 0 or at least 2. From $\frac{1}{2}\mathbb{E}[A^2]$ one gets $\sum_k m_k m_{k-1}/(k - 1)!$, of which only $k \geq 3$ survives, giving $\alpha^2(1 - \alpha)^2 s^{2k-1}\epsilon_k \epsilon_{k-1}/(k - 1)! = u^2 s \cdot s^{2k-6}\epsilon_k\epsilon_{k-1}/(k - 1)!$. In $\frac{1}{3}\mathbb{E}[A^3]$, sort the terms by the differentiated index d. If $d = 1$, the factor becomes $m_0 = 1$, the two remaining indices are at least 2 with odd sum $2(i + j + l) - 1 \geq 5$, and the term is $u^2 s$ times $s^{2(i+j+l)-6}$ times a product of two ϵ's. If $d = 2$ the term vanishes at $w = 0$. If $d \geq 3$, three factors remain, the s-degree $2(i + j + l) - 1$ is odd and at least 7, and the term is $u^2 s$ times $d\, \alpha(1-\alpha)s^{2(i+j+l)-6}$ times a product of three ϵ's. Hence $(\partial_w F)|_{w=0} = u^2 s\, P(\alpha, s)$ with, for $|s| \leq 1$, $|P| \leq \sum_{k \geq 3} \frac{1}{(k-1)!} + \frac{1}{3}\sum_{i,j,l \geq 0} \frac{(i+l)+(i+j)+(j+l)}{i!\,j!\,l!} = (e - 2) + 2e^3 = C_P$, using $\sum_{i,j,l} i/(i!j!l!) = e^3$.

The w^2 part. Let $\gamma = \partial_w^2 F - 1$. By (b), γ and its first derivatives in (w, s) vanish at $(w, s) = (0, 0)$ for every α, so Taylor's formula of order two along the segment from $(0, 0)$ to (w, s), which stays in $\mathcal{C}$, gives $|\gamma(\alpha, w, s)| \leq C_G(|w| + |s|)^2$ on $\mathcal{C}$, and $|G - \frac{1}{2}| = \big|\int_0^1 (1 - \tau)\gamma(\alpha, \tau w, s)\, d\tau\big| \leq C_G(|w| + |s|)^2$. The sup norms are finite because F is smooth on the compact set $\mathcal{C}$. $\square$

Lemma S10. *On* $\mathrm{I}(\delta_0)$*, for all* $\alpha \in [0, 1]$ *and all* $z \in \mathbb{R}$*,*

$$|A(z)| \;\leq\; u\,\Lambda_1(z) + |w|\,\Lambda_2(z)\,, \qquad \Lambda_1(z) = \tfrac{1}{2}(1 + |z|)^2 e^{|z|}\,, \qquad \Lambda_2(z) = (1 + |z|)\, e^{|z|}\,. \tag{S1.67}$$

Proof. Fix (α, s, z) and, for $|\omega| \leq |w|$, let $a(\omega)$ denote the value of $A(z)$ at the parameter with $t_1 = \omega + (1 - \alpha)s$, $t_2 = \omega - \alpha s$, so that $A(z) = a(w)$. Then $|t_h(\omega)| \leq |w| + |s| \leq 1$ by Eq. (S1.64). Since $\partial_\tau E_\tau(z) = (z-\tau)E_\tau(z)$ and $E_\tau(z) \leq e^{|z|}$ for $|\tau| \leq 1$, $|a'(\omega)| \leq \Lambda_2(z)$. At $\omega = 0$ the linear terms cancel: with $R(\tau) = E_\tau(z) - 1 - \tau z$, $a(0) = \big[\alpha(1-\alpha)s - (1-\alpha)\alpha s\big]z + \alpha R((1-\alpha)s) + (1-\alpha)R(-\alpha s)$, and the bracket is zero. Since $\partial_\tau^2 E_\tau = ((z - \tau)^2 - 1)E_\tau$ and $|(z - \tau)^2 - 1| \leq (1 + |z|)^2$ for $|\tau| \leq 1$, Taylor's formula gives $|R(\tau)| \leq \tau^2 \Lambda_1(z)$ for $|\tau| \leq \frac{2}{3}$, hence $|a(0)| \leq \big[\alpha(1-\alpha)^2 + (1-\alpha)\alpha^2\big]s^2\Lambda_1(z) = u\,\Lambda_1(z)$, and $|A| = |a(w)| \leq |a(0)| + |w| \sup|a'|$. $\square$

Proposition S4 (Normal form near the fiber)**.** *The constants* $C_L = \max(18C_G,\, 8C_P{+}16C_H)$, $C_u = 2\,\mathbb{E}\big[\Lambda_1(z)^4 e^{|z|/3+1/18}\big]$, $C_w = 2\,\mathbb{E}\big[\Lambda_2(z)^4 e^{|z|/3+1/18}\big]$ *and* $C_K = \max\left(9C_G + C_w,\ C_P,\ C_H + C_u/36\right)$

are finite. For every $\delta \in (0, \delta_0]$, where $\delta_0 = \frac{1}{3}$, and every $(\alpha, t_1, t_2) \in \mathrm{I}(\delta)$, that is, every $\alpha \in [0,1]$ and $|t| \leq \delta$:

$$K \;\geq\; F \;\geq\; (1 - C_L\delta^2)\, Q_0\,, \tag{S1.68}$$

$$0 \leq K - F \leq C_u\, u^4 + C_w\, w^4\,, \qquad |K - Q_0| \leq C_K\big(w^2|t|^2 + |w|\, u^2|s| + u^2 s^2\big) \leq 6\, C_K\, \delta^2(w^2 + u^2)\,. \tag{S1.69}$$

Proof. On $\mathrm{I}(\delta)$, $|w| \leq \delta$, $|s| \leq 2\delta$, $(|w| + |s|)^2 \leq 9\delta^2$. By Eq. (S1.66), $F \geq w^2(\frac{1}{2} - 9C_G\delta^2) - C_P\, u^2|w||s| + \frac{u^2}{4} - C_H u^2 s^2 \geq (1 - 18C_G\delta^2)\frac{w^2}{2} + \big(1 - (8C_P + 16C_H)\delta^2\big)\frac{u^2}{4}$, since $|w||s| \leq 2\delta^2$ and $s^2 \leq 4\delta^2$. Together with Lemma S9(a) this is Eq. (S1.68). For Eq. (S1.69), $K - F = \mathbb{E}[\psi_4(A)] \leq \frac{1}{4}\mathbb{E}\big[A^4 \max(1, (1+A)^{-1})\big]$ by Lemma S7. On $\mathrm{I}(\delta_0)$, $1 + A = \alpha E_{t_1} + (1-\alpha)E_{t_2} \geq \min_h E_{t_h}(z) \geq e^{-|z|/3 - 1/18}$, and $A^4 \leq 8\big(u^4\Lambda_1^4 + w^4\Lambda_2^4\big)$ by Lemma S10. Taking expectations gives the first bound. Next, $|K - Q_0| \leq |F - Q_0| + (K - F)$ with $F - Q_0 = w^2(G - \frac{1}{2}) + wu^2 sP + u^2 s^2 H$ by Eq. (S1.66), and $w^4 \leq w^2|t|^2$, $u^4 \leq u^2 s^4/16 \leq u^2 s^2/36$ by Eq. (S1.64). Collecting terms gives the second bound, and the third follows from $w^2|t|^2 \leq \delta^2 w^2$, $|w|u^2|s| \leq 2\delta^2 u^2$, $u^2 s^2 \leq 4\delta^2 u^2$. □

The constants C_L, C_u, C_w, C_K of Proposition S4 are far from optimal, and only their finiteness is used. For fixed α, K is smooth in t near the origin, since the t-derivatives of order k of $-\log(1+A)$ are dominated by polynomial multiples of $e^{2k|z|}$ uniformly for $|t| \leq 1$. Its Taylor polynomial of degree five at $t = 0$ is $\frac{w^2}{2} - \frac{uw^2}{2} + \frac{u^2}{4}$, the normal form displayed in Section 5.3 of the main paper. Indeed, only the moments $\mathbb{E}[A^r]$ with $r \leq 5$ contribute in degree at most five, and Lemma S9(b) accounts for $r = 2, 3$. By Eq. (S1.60), each monomial $\prod_{i<j}(t_{a_i} t_{a_j})^{k_{ij}}$ of $\mathbb{E}\prod_{i \leq r}(E_{t_{a_i}} - 1)$ corresponds to a multigraph on r vertices with k_{ij} edges $\{i,j\}$ and no isolated vertex, and has t-degree $2\sum k_{ij} \geq r$. Hence $\mathbb{E}[A^5]$ begins in degree six, and the quartic part of $\frac{1}{4}\mathbb{E}[A^4]$ comes from the three perfect matchings, $\frac{1}{4} \cdot 3m_1^4 = \frac{3}{4}w^4$, which cancels the $-\frac{3}{4}w^4$ of F. The cross term $-\frac{uw^2}{2}$ is of the size $w^2|t|^2$ admitted in Eq. (S1.69), and since w is of order $n^{-1/2}$ on the population ray it does not enter the limit.

The neighborhood of the fiber. For $\delta > 0$ let

$$I_n(\delta) = \int_{\mathrm{I}(\delta)} \varphi(\mu_0 + t_1)\, \varphi(\mu_0 + t_2)\, e^{-nK}\, d\alpha\, dt_1\, dt_2\,, \qquad \delta_1 = \min\big(\delta_0,\, (2C_L)^{-1/2}\big)\,, \tag{S1.70}$$

so that $c_\delta = 1 - C_L\delta^2 \geq \frac{1}{2}$ for $\delta \leq \delta_1$. For fixed $\alpha \in (0,1)$ change variables in the (t_1, t_2)-integral:

$$w = n^{-1/2} w'\,, \qquad s = n^{-1/4}\, v/\rho(\alpha)\,, \qquad dt_1\, dt_2 = dw\, ds = n^{-3/4}\rho(\alpha)^{-1}\, dw'\, dv\,, \tag{S1.71}$$

so that $u = n^{-1/2} v^2$ and the leading form becomes independent of α: $nQ_0 = \frac{w'^2}{2} + \frac{v^4}{4} = Q'(w', v)$ exactly. Let $D_n(\alpha) \subseteq \mathbb{R}^2$ be the image of the α-slice of $\mathrm{I}(\delta)$. By Eq. (S1.56),

$$\{|w'| \leq \tfrac{\delta}{2}\sqrt{n}\} \times \{|v| \leq \tfrac{\delta}{2}\, n^{1/4}\rho(\alpha)\} \;\subseteq\; D_n(\alpha) \;\subseteq\; \{|w'| \leq \delta\sqrt{n}\} \times \{|v| \leq 2\delta\, n^{1/4}\rho(\alpha)\}\,, \tag{S1.72}$$

the v-window shrinking with $\rho(\alpha)$ at the ends of the fiber. By Fubini's theorem,

$$n^{3/4} I_n(\delta) = \int_0^1 \frac{d\alpha}{\rho(\alpha)} \iint_{D_n(\alpha)} \Phi_n\, e^{-nK}\, dw'\, dv\,, \qquad \Phi_n = \varphi(\mu_0 + t_1)\varphi(\mu_0 + t_2)\,, \tag{S1.73}$$

with K and Φ_n evaluated at the point (α, t_1, t_2) corresponding to (α, w', v). The map $(w', v) \mapsto (-w', -v)$ corresponds to $(t_1, t_2) \mapsto (-t_1, -t_2)$ and preserves $D_n(\alpha)$ and K. On $D_n(\alpha)$ the point lies in $\mathrm{I}(\delta)$, so for $\delta \le \delta_1$ Proposition S4 gives $nK \ge c_\delta Q' \ge \frac{1}{2}Q'$, and the integrand of Eq. (S1.73), indicator included, is bounded by the n-free envelope

$$\|\varphi\|_\infty^2\, \rho(\alpha)^{-1} e^{-Q'(w',v)/2}\,, \qquad \int_0^1 \frac{d\alpha}{\rho(\alpha)} \iint e^{-cQ'}\, dw'\, dv = \pi \sqrt{\frac{2\pi}{c}}\, c^{-1/4}\, \frac{\Gamma(1/4)}{\sqrt{2}} \qquad (c>0)\,, \tag{S1.74}$$

using $\int_0^1 \rho^{-1} d\alpha = B(\frac{1}{2},\frac{1}{2}) = \pi$ and $\int e^{-cv^4/4} dv = c^{-1/4}\Gamma(\frac{1}{4})/\sqrt{2}$. The envelope is integrable up to the ends of the fiber because Eq. (S1.68) holds on the closed interval $\alpha \in [0,1]$ and the fiber density ρ^{-1} has finite mass, so the α-integral runs over all of $[0,1]$.

Lemma S11. *For $a > 0$ and $\lambda > 0$, with $\rho = \sqrt{\alpha(1-\alpha)}$,*

$$\int_0^1 \frac{\min(a\rho^{-2}, 1)}{\rho}\, d\alpha \le 16\sqrt{a}\,, \qquad \int_0^1 \frac{\min\big(1, (\lambda\rho)^{-3}\big)}{\rho}\, d\alpha \le \frac{12}{\lambda}\,. \tag{S1.75}$$

Proof. By symmetry integrate over $[0,\frac{1}{2}]$, where $\alpha/2 \le \rho^2 \le \alpha$. In the first integral bound the minimum by 1 on $[0, 2a]$, giving at most $\int_0^{2a} \sqrt{2}\,\alpha^{-1/2} d\alpha = 4\sqrt{a}$, and by $a\rho^{-2}$ on $[2a, \frac{1}{2}]$, giving at most $\int_{2a}^\infty a(\alpha/2)^{-3/2} d\alpha = 4\sqrt{a}$; if $2a > \frac{1}{2}$ the half integral is at most $2 \le 4\sqrt{a}$. In the second, the interval $[0, 2\lambda^{-2}]$ gives at most $4/\lambda$, and on $[2\lambda^{-2}, \frac{1}{2}]$, $\rho^{-1}(\lambda\rho)^{-3} \le 4\lambda^{-3}\alpha^{-2}$ integrates to at most $2/\lambda$; if $2\lambda^{-2} > \frac{1}{2}$ then $\lambda < 2$ and the half integral is at most $\pi/2 \le 6/\lambda$. □

Proposition S5 (The fiber)**.** *For every $\delta \in (0, \delta_1]$ there is $C < \infty$, depending on (s_0, μ_0, δ), such that $|n^{3/4} I_n(\delta) - C_2| \le C\, n^{-1/4}$ for all $n \ge 1$, with $C_2 = \varphi_0^2 \cdot \pi \cdot \sqrt{2\pi} \cdot \Gamma(\frac{1}{4})/\sqrt{2} = \pi^{3/2}\Gamma(\frac{1}{4})\varphi_0^2$.*

Proof. With $C_2 = \varphi_0^2 \int_0^1 \rho^{-1} d\alpha \iint e^{-Q'}$, decompose Eq. (S1.73) as $n^{3/4} I_n(\delta) - C_2 = T_1 + T_2 - T_3$,

$$\begin{aligned} T_1 &= \int_0^1 \frac{d\alpha}{\rho} \iint_{D_n(\alpha)} (\Phi_n - \varphi_0^2)\, e^{-nK}\, dw'\, dv\,, \qquad T_3 = \varphi_0^2 \int_0^1 \frac{d\alpha}{\rho} \iint_{D_n(\alpha)^c} e^{-Q'}\, dw'\, dv\,, \\ T_2 &= \varphi_0^2 \int_0^1 \frac{d\alpha}{\rho} \iint_{D_n(\alpha)} \big(e^{-nK} - e^{-Q'}\big)\, dw'\, dv\,. \end{aligned} \tag{S1.76}$$

On $D_n(\alpha)$ we use $nK \ge \frac{1}{2}Q'$, the chart identities $w^2 = n^{-1}w'^2$, $u = n^{-1/2}v^2$, $s^2 = n^{-1/2}v^2/\rho^2$, and the caps $|t| \le \delta$, $|s| \le 2\delta$ of $\mathrm{I}(\delta)$. In particular, since $|t|^2 \le \min\big(2w^2 + 2s^2, \delta^2\big)$,

$$|t|^2 \;\le\; 2n^{-1}w'^2 + 2\delta^2 \min\Big(\frac{n^{-1/2}v^2}{\delta^2\rho^2}, 1\Big) \qquad \text{on } D_n(\alpha)\,. \tag{S1.77}$$

T_3. By Eq. (S1.72), $D_n(\alpha)^c \subseteq \{|w'| > \frac{\delta}{2}\sqrt{n}\} \cup \{|v| > \frac{\delta}{2} n^{1/4}\rho\}$. With $I_4 = \int e^{-v^4/4} dv = \Gamma(\frac{1}{4})/\sqrt{2}$ and $\int_{|w'|>L} e^{-w'^2/2} dw' \le \sqrt{2\pi}\, e^{-L^2/2}$, the first set contributes at most $\varphi_0^2\, \pi \sqrt{2\pi}\, I_4\, e^{-\delta^2 n/8}$. Since $\int_{|v|>L} e^{-v^4/4} dv \le \min\big(I_4,\, 2\int_L^\infty (v/L)^3 e^{-v^4/4} dv\big) \le \min(I_4, 2L^{-3}) \le I_4 \min(1, L^{-3})$, as $I_4 > 2$, the second contributes at most $\varphi_0^2 \sqrt{2\pi}\, I_4 \int_0^1 \rho^{-1} \min\big(1, (\frac{\delta}{2} n^{1/4}\rho)^{-3}\big) d\alpha \le 24\, \varphi_0^2 \sqrt{2\pi}\, I_4\, \delta^{-1} n^{-1/4}$ by Lemma S11.

T_1. Let $M_2 = \max_{|\beta|=2} \sup_{\mathbb{R}^2} |\partial^\beta(\varphi(\mu_0 + t_1)\varphi(\mu_0 + t_2))| < \infty$. Then $\Phi_n - \varphi_0^2 = \varphi_0 \varphi'(\mu_0)(t_1 + t_2) + R_\varphi$ with $|R_\varphi| \le \frac{1}{2} M_2 (|t_1| + |t_2|)^2 \le 2M_2|t|^2$. The involution $(w', v) \mapsto (-w', -v)$ preserves $D_n(\alpha)$ and K and reverses the sign of $t_1 + t_2$, so the linear term integrates to zero over $D_n(\alpha)$ for every α. By Eq. (S1.77) and Lemma S11,

$$\begin{aligned} |T_1| \;&\le\; 2M_2 \iint e^{-Q'/2}\Big[\int_0^1 \frac{\mathbf{1}_{D_n(\alpha)}|t|^2}{\rho}\,d\alpha\Big]dw'\,dv \\ &\le\; 2M_2 \iint e^{-Q'/2}\big[2\pi n^{-1}w'^2 + 32\,\delta\,n^{-1/4}|v|\big]dw'\,dv \;\le\; C\,n^{-1/4}\,. \end{aligned} \tag{S1.78}$$

T_2. For $a, b \ge 0$, $|e^{-a} - e^{-b}| \le |a-b|\,e^{-\min(a,b)}$. Taking $a = nK$ and $b = Q'$, both at least $\frac{1}{2}Q'$ on $D_n(\alpha)$, and using Eq. (S1.69), $|e^{-nK} - e^{-Q'}| \le n\,C_K(w^2|t|^2 + |w|u^2|s| + u^2s^2)e^{-Q'/2}$. In the chart, by Eq. (S1.77) and the caps,

$$\begin{gathered} n\,w^2|t|^2 \le 2n^{-1}w'^4 + 2\delta^2 w'^2 \min\Big(\frac{n^{-1/2}v^2}{\delta^2\rho^2}, 1\Big), \qquad n\,|w|u^2|s| = n^{-1/2}|w'||v^4||s| \le 2\delta\,n^{-1/2}|w'|v^4, \\ n\,u^2s^2 = v^4s^2 \le 4\delta^2 v^4 \min\Big(\frac{n^{-1/2}v^2}{4\delta^2\rho^2}, 1\Big). \end{gathered} \tag{S1.79}$$

Integrating first over α against $\rho^{-1}d\alpha$ and using $\int_0^1 \rho^{-1}d\alpha = \pi$ and Lemma S11, which bounds the two capped terms by $32\,\delta\,n^{-1/4}w'^2|v|$ and $32\,\delta\,n^{-1/4}|v|^5$,

$$|T_2| \;\le\; \varphi_0^2\,C_K \iint e^{-Q'/2}\,\Psi_n\,dw'\,dv\,, \qquad \Psi_n = 2\pi\big(n^{-1}w'^4 + \delta\,n^{-1/2}|w'|v^4\big) + 32\,\delta\,n^{-1/4}\big(w'^2|v| + |v|^5\big)\,, \tag{S1.80}$$

which is at most $C\,n^{-1/4}$ for $n \ge 1$, since all moments of $e^{-Q'/2}$ are finite. □

The limit $n^{3/4}I_n(\delta) \to C_2$ alone follows from Eq. (S1.74) by dominated convergence: at fixed $\alpha \in (0,1)$ and (w', v), the point enters $D_n(\alpha)$ for large n, $\Phi_n \to \varphi_0^2$, and $nK \to Q'$ by Eq. (S1.69). The bound $n^{-1/4}$ comes from the layer $\alpha(1-\alpha) \lesssim n^{-1/2}$ at the ends of the fiber, where the quartic width $n^{-1/4}\rho(\alpha)^{-1}$ of Eq. (S1.71) is no longer small. In that layer the v-window of $D_n(\alpha)$ has width of order one (the term T_3), and the remainder terms $w^2|t|^2$, u^2s^2 of Eq. (S1.69) reach their caps (Lemma S11).

The boundary strata. On $\mathrm{II}(\delta)$ the zero set is B_0, which is not compact, and a lower bound for K must hold uniformly in $|t_1| \in (\delta, \infty)$. Since $\partial_\alpha K|_{\alpha=0} = \mathbb{E}[1 - E_{t_1}] = 0$ on $t_2 = 0$, K vanishes to second order at B_0, and the lower bound is quadratic in (α, t_2) jointly. The neighborhoods of B_0 and B_1 then contribute to $e^{-nL_0}Z_2$ at order at most n^{-1}, one transverse Gaussian factor $n^{-1/2}$ for each of α and t_2 and no decay along t_1, against $n^{-3/4}$ for the fiber.

Proposition S6 (The boundary strata)**.** *For every $\delta \in (0, \frac{1}{2}]$ there is $c_{\mathrm{II}}(\delta) > 0$ such that $K \ge c_{\mathrm{II}}(\delta)\,(\alpha^2 + t_2^2)$ for all $\alpha \in [0, \frac{1}{2}]$, $|t_2| \le \delta$, $|t_1| \ge \delta$. Consequently, for all $n > 0$,*

$$\int_{\mathrm{II}(\delta)\cup\mathrm{II}'(\delta)} \varphi(\mu_0 + t_1)\varphi(\mu_0 + t_2)\,e^{-nK}\,d\alpha\,dt_1\,dt_2 \;\le\; \frac{\pi\,\|\varphi\|_\infty}{n\,c_{\mathrm{II}}(\delta)}\,. \tag{S1.81}$$

Proof. By Eq. (S1.59) and Lemma S7, for every measurable $J \subseteq \mathbb{R}$,

$$K = \mathbb{E}[\psi(A)] \;\ge\; \frac{1}{2}\,\mathbb{E}\Big[\frac{A^2\,\mathbf{1}_J(z)}{1 + A_+}\Big]\,, \tag{S1.82}$$

and we take $J = \{|z| \le M\}$ with $M \ge 1$ chosen below. For $|\tau| \le 1$, $E_\tau(z) \le e^{|\tau||z|}$ and, from $|e^x - 1| \le |x|e^{|x|}$ with $x = \tau z - \tau^2/2$,

$$|E_\tau(z) - 1| \le |\tau|\,(1 + |z|)\,e^{|z|+1/2}\,. \tag{S1.83}$$

Far part. Let $\xi(M) = 2\,\mathbb{E}\big[(1 + (1+|z|)^2 e^{2|z|+1})\mathbf{1}_{|z|>M}\big]$, which tends to zero as $M \to \infty$. Fix $M_0 \ge 1$ with $\xi(M_0) \le \frac{1}{8}$ and put $R_0 = 4M_0$. Let $\alpha \le \frac{1}{2}$, $|t_2| \le 1$, $|t_1| \ge R_0$ and $J = \{|z| \le M_0\}$. On J, $|t_1||z| \le t_1^2/4$, so $E_{t_1} \le e^{-t_1^2/4} \le e^{-4M_0^2} = \eta$, and $A \le (1-\alpha)(E_{t_2} - 1) \le e^{M_0} - 1$, so $1 + A_+ \le e^{M_0}$. Write $A = B + \alpha E_{t_1}$ with $B = -\alpha + (1-\alpha)(E_{t_2} - 1)$, so that $A^2 \ge \frac{1}{2}B^2 - \alpha^2\eta^2$ on J. Since $\mathbb{E}[E_{t_2} - 1] = 0$ and $\mathbb{E}[(E_{t_2} - 1)^2] = e^{t_2^2} - 1$,

$$\mathbb{E}[B^2] = \alpha^2 + (1-\alpha)^2(e^{t_2^2} - 1) \;\ge\; \alpha^2 + \tfrac{1}{4}t_2^2 \;\ge\; \tfrac{1}{4}(\alpha^2 + t_2^2)\,, \tag{S1.84}$$

and by Eq. (S1.83) $B^2 \le 2\alpha^2 + 2t_2^2(1+|z|)^2 e^{2|z|+1}$, so $\mathbb{E}[B^2\mathbf{1}_{J^c}] \le (\alpha^2 + t_2^2)\xi(M_0) \le \frac{1}{8}(\alpha^2 + t_2^2)$. Hence $\mathbb{E}[A^2\mathbf{1}_J] \ge \frac{1}{2}\big(\frac{1}{4} - \frac{1}{8}\big)(\alpha^2 + t_2^2) - \alpha^2\eta^2 \ge \big(\frac{1}{16} - e^{-8}\big)(\alpha^2 + t_2^2) \ge \frac{1}{32}(\alpha^2 + t_2^2)$, and by Eq. (S1.82) $K \ge c_{\mathrm{far}}(\alpha^2 + t_2^2)$ with the absolute constant $c_{\mathrm{far}} = e^{-M_0}/64$.

Near B_0, bounded t_1. The function $V = \mathbb{E}[A^2] = \alpha^2(e^{t_1^2} - 1) + 2\alpha(1-\alpha)(e^{t_1 t_2} - 1) + (1-\alpha)^2(e^{t_2^2} - 1)$ is entire. At $(\alpha, t_2) = (0,0)$, for every t_1, $V = \partial_\alpha V = \partial_{t_2} V = 0$, and the Hessian in (α, t_2) is $2\mathsf{H}(t_1)$ with

$$\mathsf{H}(t_1) = \begin{pmatrix} e^{t_1^2} - 1 & t_1 \\ t_1 & 1 \end{pmatrix}, \qquad \det \mathsf{H}(t_1) = e^{t_1^2} - 1 - t_1^2 > 0 \quad (t_1 \ne 0)\,. \tag{S1.85}$$

Let $\lambda_\delta > 0$ be the minimum over $\delta \le |t_1| \le R_0$ of the smaller eigenvalue of $\mathsf{H}(t_1)$, and M_3 the maximum of the absolute values of the third-order partial derivatives of V in (α, t_2) over $\{\alpha \in [0, \frac{1}{2}],\ |t_2| \le 1,\ |t_1| \le R_0\}$. With $r^2 = \alpha^2 + t_2^2$, Taylor's theorem at $(\alpha, t_2) = (0,0)$ gives $V \ge \lambda_\delta r^2 - C_V r^3$ for $\alpha \in [0, \frac{1}{2}]$, $|t_2| \le 1$, $\delta \le |t_1| \le R_0$, with $C_V = 2^{3/2}M_3/6$. On $J = \{|z| \le M\}$, $1 + A_+ \le e^{R_0 M}$, and by Eq. (S1.83) $A^2 \le 2\alpha^2(e^{R_0|z|} + 1)^2 + 2t_2^2(1+|z|)^2 e^{2|z|+1}$, so $\mathbb{E}[A^2\mathbf{1}_{J^c}] \le r^2\xi_1(M)$ with $\xi_1(M) = 2\,\mathbb{E}\big[((e^{R_0|z|} + 1)^2 + (1+|z|)^2 e^{2|z|+1})\mathbf{1}_{|z|>M}\big] \to 0$. Choose $M = M_1$ with $\xi_1(M_1) \le \lambda_\delta/4$ and put $r_0 = \lambda_\delta/(4C_V)$. For $r \le r_0$, $\mathbb{E}[A^2\mathbf{1}_J] \ge \lambda_\delta r^2 - C_V r^3 - \frac{\lambda_\delta}{4}r^2 \ge \frac{\lambda_\delta}{2}r^2$, and Eq. (S1.82) gives $K \ge c_{\mathrm{near}}(\delta)\,r^2$ with $c_{\mathrm{near}} = \lambda_\delta e^{-R_0 M_1}/4$ on $\{\alpha \le \frac{1}{2},\ |t_2| \le 1,\ \delta \le |t_1| \le R_0,\ r \le r_0\}$.

The rest. The set $S = \{\alpha \in [0, \frac{1}{2}],\ |t_2| \le \delta,\ \delta \le |t_1| \le R_0,\ r \ge r_0\}$ is compact and disjoint from $\mathcal{Z}$ ($\mathcal{R}$ and B_1 are excluded by $|t_1| \ge \delta$ and $\alpha \le \frac{1}{2}$, B_0 by $r \ge r_0$), so $\kappa_S = \min_S K > 0$ by Lemma S8(b),(c), and $K \ge 2\kappa_S r^2$ on S since $r^2 \le \frac{1}{4} + \delta^2 \le \frac{1}{2}$ there.

Since $\delta \le \frac{1}{2} < R_0$, the three cases cover the set $\alpha \le \frac{1}{2}$, $|t_2| \le \delta$, $|t_1| \ge \delta$, and the bound holds with $c_{\mathrm{II}}(\delta) = \min\big(c_{\mathrm{far}},\, c_{\mathrm{near}}(\delta),\, 2\kappa_S\big)$. For Eq. (S1.81), bound $\varphi(\mu_0 + t_2) \le \|\varphi\|_\infty$, integrate $\varphi(\mu_0 + t_1)$ to one, and use $\int_0^\infty\!\int_{\mathbb{R}} e^{-nc(\alpha^2 + t_2^2)}dt_2\,d\alpha = \pi/(2nc)$. The involution $(\alpha, t_1, t_2) \mapsto (1-\alpha, t_2, t_1)$ maps II' into the set where the bound holds and preserves K and the prior. $\square$

Proposition S7 (Away from the zero set)**.** *For every $\delta \in (0, \frac{1}{2}]$, $\kappa(\delta) = \inf_{\mathrm{III}(\delta)} K > 0$, and hence $\int_{\mathrm{III}(\delta)} \varphi(\mu_0 + t_1)\varphi(\mu_0 + t_2)e^{-nK}\,d\alpha\,dt_1\,dt_2 \le e^{-n\kappa(\delta)}$.*

Proof. Choose $T \ge 1$ with $e^{-T^2/4} \le \frac{1}{2}(1 - e^{-\delta^2/4})$ and $e^{-T^2/8} \le \frac{1}{2}(1 - 2^{-1/2})$. On $\mathrm{III}(\delta) \cap \{|t| \ge T\}$: if both $|t_h| > \delta$, then one of them is at least T and $b^2 \le e^{-T^2/4} + e^{-\delta^2/4} \le 1 - \frac{1}{2}(1 - e^{-\delta^2/4}) = b_1^2 < 1$ by Lemma S8(d); if $\alpha > \frac{1}{2}$, $|t_1| > \delta$, $|t_2| \le \delta$, then $|t_1| \ge T$ and $b \le e^{-T^2/8} + \sqrt{1-\alpha} \le \frac{1}{2}(1 + 2^{-1/2}) = b_2 < 1$; the third piece is symmetric. Hence $K \ge -2\log\max(b_1, b_2) > 0$ there. The part of $\mathrm{III}(\delta)$ with $|t| \le T$ is contained in the union of the compact sets $\{\delta \le |t_1| \le T,\ \delta \le |t_2| \le T\}$,

$\{\alpha \geq \frac{1}{2},\ \delta \leq |t_1| \leq T,\ |t_2| \leq \delta\}$, $\{\alpha \leq \frac{1}{2},\ \delta \leq |t_2| \leq T,\ |t_1| \leq \delta\}$, none of which meets $\mathcal{Z}$: on $\mathcal{R}$, $t_1 = t_2 = 0$; on B_0, $\alpha = 0$ and $t_2 = 0$; on B_1, $\alpha = 1$ and $t_1 = 0$. By Lemma S8(b),(c), K has a positive minimum on each. The integral bound follows since the prior has mass one. $\square$

Proof of Proposition S3. (i) $L_1(\mu) = \int q \log N(y;\mu,1)\,dy = L_0 - (\mu-\mu_0)^2/2$, so $e^{-nL_0} Z_1(n) = \int \varphi(\mu)\, e^{-n(\mu-\mu_0)^2/2} d\mu = \sqrt{2\pi/n} \int N(\mu; 0, s_0^2)\, N(\mu_0; \mu, \frac{1}{n})\, d\mu = \sqrt{2\pi/n}\; N(\mu_0; 0, s_0^2 + \frac{1}{n})$, which is the closed form, and its ratio to $\varphi_0\sqrt{2\pi/n}$ is $(1 + \frac{1}{n s_0^2})^{-1/2} \exp\{\mu_0^2/(2s_0^2(1+ns_0^2))\} = 1 + \mathcal{O}(1/n)$.

(ii) By Eq. (S1.54) and the partition of $[0,1]\times\mathbb{R}^2$ into $\mathrm{I}, \mathrm{II}, \mathrm{II}', \mathrm{III}$ at $\delta = \delta_1 \leq \frac{1}{3}$, $e^{-nL_0} Z_2(n) = I_n(\delta_1) + \int_{\mathrm{II}\cup\mathrm{II}'} + \int_{\mathrm{III}}$ with the same integrand. By Propositions S5, S6 and S7, $|n^{3/4} e^{-nL_0} Z_2(n) - C_2| \leq C\, n^{-1/4} + \pi \|\varphi\|_\infty\, n^{-1/4}/c_{\mathrm{II}}(\delta_1) + n^{3/4} e^{-n\kappa(\delta_1)} = \mathcal{O}(n^{-1/4})$. The two expressions for C_2 agree since $\varphi_0^2 = e^{-\mu_0^2/s_0^2}/(2\pi s_0^2)$.

(iii) Dividing (i) by (ii),

$$\mathrm{BF}(n) = \frac{\varphi_0\sqrt{2\pi}\, n^{-1/2}(1+\mathcal{O}(1/n))}{n^{-3/4}(C_2 + \mathcal{O}(n^{-1/4}))} = \frac{\varphi_0\sqrt{2\pi}}{C_2}\, n^{1/4}(1 + \mathcal{O}(n^{-1/4}))\,, \tag{S1.86}$$

and $\varphi_0\sqrt{2\pi}/(\pi^{3/2}\Gamma(\frac{1}{4})\varphi_0^2) = \sqrt{2}/(\pi\,\Gamma(\frac{1}{4})\,\varphi_0) = 2s_0\, e^{\mu_0^2/(2s_0^2)}/(\sqrt{\pi}\,\Gamma(\frac{1}{4}))$. $\square$

The sample size at which the leading term of (iii) reaches odds B is $(B/c)^4 = B^4\pi^4\Gamma(\frac{1}{4})^4\varphi_0^4/4$, the form used in Section 6.1 of the main paper. The relative correction, at most of order $n^{-1/4}$, comes from the ends of the fiber (Proposition S5) and from the boundary strata (Proposition S6), whose contributions to $e^{-nL_0} Z_2$ are both of absolute order at most n^{-1}. The estimates above bound this correction without determining its coefficient. A direct numerical quadrature of Eq. (S1.54) indicates that the coefficient is nonzero: the quantity $\big(n^{3/4} e^{-nL_0} Z_2(n)/C_2 - 1\big)\, n^{1/4}$ equals -1.07, -1.21, -1.26 at $n = 10^4$, 10^6, 10^8 for $(s_0, \mu_0) = (1, 0)$, and -0.65, -0.75, -0.78 for $(s_0,\mu_0) = (1,1)$. Every statement in this section is a population-ray statement, and the sampled-data constant remains open.

S2 Boundary-stratum computations and numerical checks

In this section we prove the χ^2-factorization (Theorem 8 of the main paper), state the family and general-size laws behind Theorems 9 and 10 of the main paper together with the computations that give them, evaluate the flat integrals, and compare the constants with the evidence computed directly at finite N.

S2.1 The Hessian entries and the factorization theorem

For naive Bayes at uniform data we use the unnormalized exponent, the unweighted sum of $\ln M_v$ over the cells, in place of the q-normalized exponent of Theorem 8 of the main paper. Write $p_v(\theta) = 2^{-F}\prod_f(1 + s_f A_f)$ with $s_f = 2v_f - 1$, $A_f = 2\theta_f - 1$; sums over v of products of the sign characters collapse by orthogonality, giving $-\partial_\sigma^2\Phi = 2^{2F}\sum_v (p_v - 2^{-F})^2 = 2^F[\prod_f(1+u_f) - 1]$ and similarly the other entries, with $u_f = (2\theta_f - 1)^2$; taking the Schur complement of the diagonal block of the second component's parameters η gives the determinant, $\det H = 2^{F(F+2)}\cdot 2^F[\prod(1+u_f) - 1 - \sum u_f]$, and the bracket is $\sum_{j\geq 2} e_j(u)$. For the tables, the block of the second component's row distribution c is the multinomial Fisher matrix (determinant $1/\prod_i \rho_i$ by Sherman–Morrison), $H_{\sigma c_i}$ is proportional to the χ-score of a against ρ, and the σ-Schur complement collapses to $\chi^2(a\|\rho)\chi^2(b\|\kappa)$. The naive Bayes and the table computations are both instances of Theorem 8, which we now prove.

Proof of Theorem 8 of the main paper. (i) Arrow form. For any regular family, on S_0 the mixture satisfies $\partial_\sigma M_v = p_v(\omega_1) - q_v$ and $M_v = q_v$, so $-\partial_\sigma^2 \Phi = \sum_v q_v\,[(p_v(\omega_1) - q_v)/q_v]^2 = \chi^2(P\|q)$ with $P = p(\omega_1)$; the (ω_2, ω_2) block of $-\partial^2\Phi$ is the Fisher information of the true family at the truth; and the mixed (σ, ω_2) entries are $g = \sum_v (p_v(\omega_1) - q_v)\,\partial_{\omega_2} \ln p_v(\omega^*)$, the χ-scores (the remaining terms vanish by the sum-to-one identity). The Hessian is therefore the bordered Gram (arrow) matrix with corner $\chi^2(P\|q)$, diagonal blocks the per-factor Fisher matrices I_f, and border g.

(ii) Product structure. For product families the corner, the Fisher block, and the border each factorize over the factors: $1 + \chi^2(P\|q) = \prod_f (1 + \chi_f^2)$, by Fubini applied to $\sum_v P_v^2/q_v = \prod_f \sum_s (P_s^{(f)})^2/q_s^{(f)}$; the Fisher matrix is block diagonal, one multinomial Fisher block per factor (mean-zero scores of distinct factors are uncorrelated under the product truth). The border g splits into per-factor pieces g_f because the score of factor f depends only on v_f, so that in g the sum over the remaining coordinates reduces $P - q$ to its f-th marginal deviation.

(iii) Saturation. Each factor family is the full multinomial on its alphabet, so its scores span the mean-zero functions on it; the regression of the deviation $P^{(f)} - q^{(f)}$ on the factor scores therefore has zero residual, and $g_f^\top I_f^{-1} g_f = \chi_f^2$. This is the only step that uses saturation: for a curved factor family the scores span a proper subspace, the regression identity fails, and so does the factorization.

(iv) Schur complement. By the arrow structure,

$$\det H \;=\; \det I \cdot \Big[\chi^2(P\|q) - \sum_f g_f^\top I_f^{-1} g_f\Big] \;=\; \det I \cdot \Big[\prod_f (1 + \chi_f^2) - 1 - \sum_f \chi_f^2\Big] \;=\; \det I \sum_{j \ge 2} e_j(\chi^2)\,, \tag{S2.1}$$

and Sherman–Morrison gives $\det I_f = 1/\prod_s q_s^{(f)}$ in the reduced per-factor coordinates. □

For F binary factors, $\chi_f^2 = w_f = (\theta_f - \theta_f^*)^2/[\theta_f^*(1 - \theta_f^*)]$; at the uniform point $\chi_f^2 = u_f = (2\theta_f - 1)^2$, and for the unnormalized exponent used above the prefactor of $\sum_{j\ge 2} e_j(u)$ in $\det H$ is $2^{F(F+3)}$. For $F = 2$ factors of size k, $\sum_{j \ge 2} e_j = e_2 = \chi^2(a\|\rho)\,\chi^2(b\|\kappa)$, and at the uniform point the unnormalized $\det H$ is $k^{6k}|a - u|^2|b - u|^2$. The factorization of Theorem 8 holds in the same form when the factors have different alphabet sizes, for example in a $2 \times 2 \times 3$ model, which belongs to neither family above.

S2.2 The family and general-size laws

Theorem S1 (Naive Bayes family on the type-2 stratum)**.** *At any interior product data $q_v = \prod_f (\theta_f^*)^{v_f}(1 - \theta_f^*)^{1 - v_f}$ with $F = 3$, $Z_N = C(\theta^*) N^{-2} e^{N \sum_v q_v \ln q_v}(1 + o(1))$ with $C(\theta^*)$ as displayed in Section 5.4 of the main paper.*

Theorem S1 follows from the factorization of Theorem 8 written in the χ^2 coordinates w_f; the octants of the flat cube around θ^* then contribute incomplete-E_1 factors whose arguments are the odds ratios displayed in Section 5.4 of the main paper.

Proposition S8 (General $F \ge 3$)**.** *The Hessian law holds for every F, giving $Z_N = C_F\, N^{-(F+1)/2}\, 2^{-FN}(1 + o(1))$ with*

$$C_F = (2\pi)^{(F+1)/2}\, 2^{-F} J_F\,, \qquad J_F = \int_{[1,\infty)^F} \prod_f \frac{dx_f}{x_f} \Big(\sum_{i=0}^{F-2} e_i(x^2)\Big)^{-1/2}, \tag{S2.2}$$

where $\lambda = (F+1)/2$ *itself is the theorem of Rusakov and Geiger (2005);* $J_3 = J$*, and* $J_4 = 2.117832\ldots$ *by quadrature of the integral.*

The integral J of Theorem 9 of the main paper is evaluated by three substitutions: $t_f = |2\theta_f - 1|$ maps the flat integral onto the unit cube; under the reciprocal substitution $x_f = 1/t_f$ one has $e_2(t^2) + e_3(t^2) = (t_1t_2t_3)^2(1+|x|^2)$, so that the flat integral becomes $\int_{[1,\infty)^3} \prod_f (dx_f/x_f)\,(1+|x|^2)^{-1/2} = J$; and the Schwinger representation of the inverse square root factorizes the triple integral, each x_f-integral giving one factor of E_1.

Theorem S2 (Table family on Lin's stratum)**.** *At any positive rank-one data* $q = \rho \otimes \kappa$ *on* 3×3 *tables,*

$$Z_N = C(\rho,\kappa)\, N^{-5/2}\, e^{N\sum q_{ij}\ln q_{ij}}(1+o(1))\,, \qquad C(\rho,\kappa) = (2\pi)^{5/2}\sqrt{\prod_i \rho_i \prod_j \kappa_j}\; I(\rho)\, I(\kappa)\,, \quad \text{(S2.3)}$$

with $I(\rho) = \int_{\Delta_2} da/\sqrt{\chi^2(a\|\rho)}$ *elementary for every* ρ*;* $I(u) = \sqrt{2}\ln(2+\sqrt{3})/\sqrt{3}$ *recovers Theorem 10 of the main paper.*

Proposition S9 (General $k \geq 3$)**.** *The boundary strata contribute* $C_k N^{-(2k-1)/2} k^{-2N}$ *with*

$$C_k = (2\pi)^{(2k-1)/2}\, k^{-(k+1)}\, I_1(k)^2\,, \qquad I_1(k) = \int_{\Delta_{k-1}} \frac{da}{|a-u|}\,, \tag{S2.4}$$

and dominate for every $k \geq 3$*: the critical variety at the uniform ray consists of the two boundary strata and the rank-one sheets meeting in the center; every non-boundary stratum enters at* $O(N^{-3(k-1)/2}\ln N)$*, a gap of* $(k-2)/2$ *powers, and an elementary Pinsker–volume argument gives the log-free bound* $Z_N e^{NE} \leq B_k N^{-(2k-1)/2}$ *for all* N*. Hence* $\lambda = (2k-1)/2$ *and* $m = 1$ *for every* $k \geq 3$*, recovering the theorem of Lin (2017) at* $k = 3$*. At* $k = 2$ *the model is regular-saturated and the analysis is void.*

For the naive Bayes model at uniform data the strata contribute at orders N^{-2} (boundary), then $N^{-9/4}$ (the type-1 segments carrying the quartic degeneracy), then $N^{-5/2}$ (interior sheets); for the tables the orders are $N^{-5/2}$ (boundary), then N^{-3} (margin families), then $N^{-3}\log N$ (the center segment, whose hyperbolic quartic produces the logarithm).

The triangle integral. In centered coordinates $|a-u|^2 = 2(x^2+xy+y^2)$, and by the threefold symmetry $I_1 = 3\int_S$ with S the sector subtended by the bottom edge. In polar coordinates the sector integral is $\int R\, g^{-1/2} d\vartheta$ with $g = 2 + \sin 2\vartheta$, $R = -1/(3\sin\vartheta)$, over $\vartheta \in (-3\pi/4, -\arctan\frac{1}{2})$. With $t = \tan\vartheta$ (splitting at $-\pi/2$) both pieces reduce to $\frac{1}{3\sqrt{2}}\int dt/(t\sqrt{t^2 \pm t + 1})$, and with

$$\int \frac{dt}{t\sqrt{t^2+bt+1}} \;=\; -\ln\frac{2+bt+2\sqrt{t^2+bt+1}}{t} \tag{S2.5}$$

the four endpoint evaluations combine to $\frac{1}{3\sqrt{2}}[\ln(1+\frac{2}{\sqrt{3}}) + \ln(3+2\sqrt{3})] = \frac{\sqrt{2}}{3}\ln(2+\sqrt{3})$, since $(1+2/\sqrt{3})(3+2\sqrt{3}) = (2+\sqrt{3})^2$. The three sectors together give $I_1 = \sqrt{2}\ln(2+\sqrt{3})$, the flat integral of Theorem 10 of the main paper. For general ρ the same computation gives $I(\rho)$: the quadratic form becomes $\sum_i x_i^2/\rho_i$ with $x = a - \rho$, and the three sectors are delimited by the directions of the triangle vertices as seen from ρ.

S2.3 Localization and numerical evaluation

The Laplace localization used in the theorems of this section follows from standard estimates: the contributions of the different strata add, and the degenerate corners of the flat integrals (where $\det H \to 0$; all flat integrals converge there) and the interior strata enter at the correction orders stated above. A dominated-convergence argument with an N-free integrable envelope, valid up to the degenerate corners precisely because $F \geq 3$, resp. $k \geq 3$, proves the localization uniformly, with no splitting parameter; an N-dependent two-zone estimate with splitting radius $N^{-2/5}$ gives the same limits. For the table family the dominance of the boundary strata does not rest on these estimates: the log-free global bound of Proposition S9 controls every non-boundary contribution unconditionally, and the standard estimates are needed only for the $o(1)$ in the constant.

We evaluated the evidence numerically at five data rays: uniform and $\theta^* = (\frac{3}{4}, \frac{1}{2}, \frac{1}{2})$ for naive Bayes; uniform, $\rho = \kappa = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, and $\rho = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $\kappa = (\frac{3}{4}, \frac{1}{8}, \frac{1}{8})$ for tables. At every ray the evidence at small N was computed as an exact rational, both by a big-integer dynamic program over sufficient statistics and by term-by-term monomial expansion of the parameter integral. The floating-point evaluations agree with these exact values at all comparison points, to 10^{-10} at the uniform rays, 3×10^{-8} at the generic naive Bayes ray, and 10^{-7} (single precision) at the generic table rays. Running the same dynamic program in log-space floating point, we computed the evidence up to $N = 384$ for naive Bayes and $N = 192$–216 for the tables. With $y_N = \ln Z_N + NE$ and n the per-cell count where applicable, the computed sequences are as follows. Uniform naive Bayes ($N = 8n$): $y = -5.48431, -6.22232, -6.75554, -7.17359, -7.51763, -7.81005, -8.06437, -8.49129, -8.84168$ at $n = 8, 12, 16, 20, 24, 28, 32, 40, 48$ (double precision; exact rational values at $n \leq 4$). Naive Bayes at $\theta^* = (\frac{3}{4}, \frac{1}{2}, \frac{1}{2})$, $N = 32$–384: -4.56669, -5.78083, -6.51872, -7.05146, -7.81259, -8.35857, -8.78487, -9.13478. Uniform tables ($N = 9n$), $n = 8, \ldots, 24$: -9.21258, -10.15135, -10.83288, -11.36861, -11.81021. Tables at $\rho = \kappa = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $N = 32$–192: -7.78666, -8.65400, -9.29398, -10.22375, -10.89942, -11.43096, -11.86937. Tables at $\rho = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $\kappa = (\frac{3}{4}, \frac{1}{8}, \frac{1}{8})$, $N = 32$–192: -8.73919, -10.27138, -11.20354, -11.87785, -12.40721, -12.84336. The exact rational value at $N = 32$ on the last ray is $y = -8.739193763$, and at the two largest values of N computed in both single and double precision the two evaluations agree to 2.1×10^{-7} and 4.7×10^{-7}.

With the leading $-\lambda \ln N$ term of y_N removed, the naive Bayes sequences approach their constants to within a few percent with $N^{-1/2}$ corrections. The exponent estimates obtained from consecutive differences of y_N increase monotonically toward 2 for naive Bayes (reaching 1.92) and toward 5/2 for the tables (reaching 2.37–2.42). However, the table sequences have not entered the asymptotic regime by $N = 216$: the residual in $\ln Z$ is $+0.18$ there and still rising, so for the tables the constant rests on the localization proof. A center-stratum correction of the form $(c_1 \ln N + c_2)N^{-1/2}$, whose maximum lies beyond the feasible range of N, is consistent with the computed values. Fits of that form with the intercept fixed at the theoretical constant reproduce the table sequences to rms 8×10^{-7} (uniform), 4×10^{-7} (first generic ray) and 5×10^{-5} (second generic ray), although such a fit does not confirm the constant numerically. At the uniform table ray a fit on the window $n = 4$–20 predicted the next computed value to 6×10^{-6}, well inside the fit's own uncertainty of order 10^{-4}.

S3 Numerical methods and holonomic structure

The exact evidence values in the main paper are computed by integer dynamic programming over the sufficient-statistic lattice, and the floating-point values by log-stabilized or exponentially tilted evaluations of the same sums, as described below. The evidence integrals are also holonomic in the data counts; for the binary model we write the contiguity relations out, whereas for the binomial mixture we find no scalar recurrence within the orders and degrees we could search. The numerical checks behind each computation are collected in the last subsection, keyed to the sections of the main paper.

S3.1 Numerical methods

The dynamic program for the rational evidence values runs in exact integer arithmetic for $N \leq 96$ and, for the larger benchmarks, modulo 25-bit primes, followed by Chinese remaindering and rational reconstruction. The reconstruction is confirmed at additional primes not used in it; for evidence integrals the numerator and denominator sizes are highly asymmetric, and the reconstruction bounds must be chosen accordingly. The large-N values in Table 2 of the main paper use a log-stabilized dynamic program for the allocation sum: each lattice row carries a floating-point profile and a per-row logarithmic offset, so the full dynamic range, spanning thousands of orders of magnitude between the ridge and the corners, is handled without underflow. This evaluator agrees with the exact rational values to at least 13 significant figures at every point where both were computed ($N \leq 96$ for the binomial mixture, $N \leq 600$ for the binary model against Theorem 1 of the main paper). Its convergence was confirmed by recomputation with more working digits at selected N. The floating-point rows of Table 1 of the main paper use exponentially tilted arithmetic: the generating-function coefficients are rescaled by a geometric weight chosen so that the dominant terms are of order one, and the weight is removed in logarithmic form at the end.

S3.2 Holonomic structure and a computational negative result

The evidence integrals of the main paper are Euler-type integrals with the data counts in the exponents, so $Z(U)$ is holonomic in U: it satisfies contiguity recurrences with polynomial coefficients, count-shift relations in the hypergeometric sense. Recurrences of this kind have been developed for the conditional normalizing constants of contingency tables (Tachibana et al., 2020), and the A-hypergeometric structure has been worked out for partition distributions on the rational normal curve, the variety underlying the binomial mixture (Mano, 2017), though never, to our knowledge, for evidence integrals of latent-variable models.

For the binary model we derived the contiguity system explicitly. The evidence satisfies the sum rule $Z(U) = Z(U + e_0) + Z(U + e_1)$ and an order-three recurrence in a single count direction,

$$c_0(u_0, u_1)Z(U) + c_1 Z(U + e_0) + c_2 Z(U + 2e_0) + c_3 Z(U + 3e_0) \;=\; 0\,, \tag{S3.1}$$

with coefficients of degree three, e.g. $c_0 = -(u_0 + 1)(u_0 + 2)^2$; iterating this recurrence upward from small counts reproduces the exact rational values of Theorem 1 of the main paper. We also constructed an integration-by-parts calculus for the parameter integrals using vector fields $w(1 - w)\partial_w$ tangent to the simplex boundaries, which generates recurrence systems without boundary terms. We checked this calculus in full on the binary model: the recurrence system it produces there is satisfied by the closed form of Theorem 1.

For the binomial mixture we looked for a scalar recurrence in n along its singular data ray $U = n\,(1, 4, 6, 4, 1)$ (Section 5.1 of the main paper) and found none. If one exists at the orders we examined, its coefficient degree lies above every cap we could test, and these caps are as large as 140 at low order. We computed the exact series $Z(n)$ modulo fourteen 25-bit primes to $n = 800$ (any recurrence proposed for this family must annihilate these residue sequences) and ran overdetermined recurrence searches on them, each repeated at a second prime with the same outcome. An 800-term series bounds the search to the region $(r+1)(d+1) \lesssim 670$ in recurrence order r and coefficient degree d: the tested per-order degree caps decrease from 140 at $r \leq 3$ to 37 at $r = 16$. Within this region no operator annihilates the sequence on the singular ray, the sequence on the all-ones ray, or, at recurrence orders $r \leq 10$, either of the two Gamma-factor renormalizations $Z_n\,(16n+1)!$ and $Z_n\,(16n+1)!\,(32n+1)!$. Diagonal restrictions of multivariate holonomic sequences can have far larger scalar complexity than the underlying multivariate systems (Lipshitz, 1988; Bostan et al., 2017), and this family appears to be a strong instance. We therefore did not pursue a scalar recurrence for the large-n asymptotics, and none is needed: in Section 5 of the main paper the learning coefficient λ and the constant term are obtained directly from the allocation sum by localizing it to a tube around its ridge.

S3.3 Numerical checks for the main-text computations

Section, table and theorem numbers here refer to the main paper.

Sections 2–4. Our implementation reproduces the three published values of Lin et al. (2009) named in Section 2.1 to every printed digit, including the coin-toss fraction with its 530-digit numerator and 552-digit denominator and, by modular evaluation at 64 primes, the fraction of their Example 5.5 with a 143-digit numerator and 262-digit denominator. In exact arithmetic we checked Theorem 1 against the allocation sum for all $0 \leq u_0, u_1 \leq 8$, at $(25, 25)$, and against symbolic integration at six count pairs, and Theorem 4 against brute-force allocation sums at $g = 2, 3, 4$ (zero counts included) and symbolic integration. The same checks cover its $\mathrm{Dir}(\beta_0)$ variant at five rational β_0, the coefficient $-1/(48g)$ of Section 4.3 at $g = 64, 128, 256$, and the reductions of Appendix A at every integer point of each polynomial piece. In Table 1, four exact rows were reproduced by a second implementation. The tilted floating-point rows agree with exact evaluation to relative error 3.1×10^{-12} or better at $N = 100$ for $g = 10^2, 10^4, 10^6$ and to 7.5×10^{-10} in $\ln Z$ at $g = 10^6$, $N = 5{,}000$. The $\alpha = 1$ Dirichlet-process rows agree with exact partition sums to 2.7×10^{-15} in $\ln Z$, the $\alpha = 2$ row with exact-rational evaluation to 6.3×10^{-16} (checked at $N = 80$), and the $\alpha = 1$ rows also with quadrature against the Diaconis–Kemperman density to 10^{-9} at $N = 5{,}000$ and to every printed decimal at $N = 50{,}000$. The times in Table 1 come from more than one machine. On a single machine the $\alpha = 1$ row at $N = 5{,}000$ took 21 s against 27 s for $g = 10^6$, so the printed times should not be used to rank the cost of the Dirichlet-process and large-g evaluations.

Section 5. The identities in Lemmas 1–2, as well as the normal form and E_1 reduction of Section 5.3, were verified by computer algebra. The Hessian entries, the vanishing of the gradient along the strata and the determinant identities of Section 5.4 were confirmed symbolically through $F = 6$ for naive Bayes and $k = 5$ for the $k \times k$ tables. For Theorem 7 at $\theta_0 = 1/3$, a two-coefficient correction fit on $N = 648, 972$ predicts the computed values at $N = 1296, 1944$ to 2×10^{-4} in $\ln Z$. Our values reproduce the printed F_N differences of Lin et al. (2009) to 2×10^{-8}. In Table 3 the fitted exponents agree with $1/4$ within 1.3×10^{-4}, and for reduced-rank regression the fit recovers both constants of the asymptotic expansion within uncertainties of 1.8×10^{-11} and 7.1×10^{-10}; neither fit uses the closed form. The constant $J_4 = 2.117832$ is a quadrature value not yet compared with evidence evaluations, and integer-relation searches at fifty digits for the three-feature constant J found no closed form. Quadratic decay of the interior strata (Section 5.4) was checked by second-difference ratio tests.

Section 6.1. Exact Bayes factors at $N = 400$, 800, 1600 are 2.288, 2.688, 3.161, against 2.067, 2.458, 2.923 from the asymptotic law. Direct quadrature (relative error below 4×10^{-11}) gives BF(1,281) = 3.00006, and the values BF = 19.977 and 20.051 at $N = 3.30$ and 3.35×10^6 bracket the strong-evidence crossing. The formulas of Section 6.1 giving the Bayes-factor constant and the required sample size as functions of the trial design were re-derived symbolically for each m and checked numerically at nine m from 3 to 20. The BIC crossing at $N = 22$ clears the cutoff by at least 6×10^{-4} in deviance, against optimization error below 10^{-6}.

Section 6.2. The estimator comparison used 52,010 runs against 1,080 reference values (630 exact rationals, 270 quadratures agreeing at two resolutions to 2×10^{-10}, 180 log-gamma evaluations), and 9,825 further runs for the two permutation variants. At $N = 10^6$ the permutation-averaged Chib estimator recovers the $\log 2$ (+0.6931 nats) that the raw Chib estimate misses on the binary family and lowers the binomial error from 1.449×10^6 to 9.166×10^5 nats. In two sets of deliberately perturbed reference values inserted as positive controls, every perturbation was detected at 3σ by the estimators tested, apart from the harmonic mean in the first set (excluded beforehand as ill-behaved) and the raw Chib estimator in the second ($|z| = 2.90$ against the 3.0 threshold). A stepping-stone re-run under a later version of the code differed from the original runs by 2.7×10^{-2} nats (1.3 times the estimator's error estimate), a shift whose cause we have not identified; Table 4 uses the original values. In Table 4 the number of completed nested-sampling runs on the binomial family is 250 because twenty runs at $N \geq 3 \times 10^5$ exceeded the time limit of the computing queue. The error bar of the permutation-averaged Chib row is that of the raw Chib row, and the flip count of bridge sampling on the binomial family was not recorded. For the coarsened posterior, the 0.193 slope and the $\alpha = 10^2$ half-height are fits through medians, whereas the $\alpha = 10^3$ value and, for $N \leq 10^4$, the $\alpha = 10^4$ value are supported by interval enclosures of relative radius at most 3.5×10^{-7} at 209 points. At 202 of these a Laplace approximation taken over one half of the label-symmetric parameter space and doubled lies 0.511 nats low at the median, inside the $[0, \ln 2]$ band expected of that construction when the posterior mode straddles the symmetry plane. The record-linkage Gibbs sampler uses the same likelihood tables as the exact reference posteriors and was first compared with full enumeration on three small blocks (200,000 sweeps). Over the 24 configurations the worst total-variation distance was 0.05141 (0.98–1.45 times the iid Monte Carlo floor) and the worst mean shift 0.0419 posterior standard deviations, against preset limits of 0.1 and 2.

Sections 6.3 and 7.1. The eyetracking counts in the original printed table agree with both copies of the data distributed in statistical software packages. The α-free inner sums took 39 s for the two-cell binning and 22 min for the three-cell binning and were checked against partition enumeration on small subsets; at $\alpha = 1$, $k = 2$ the evidence is a fraction with a 180-digit numerator and a 211-digit denominator. In Figure 2 the mass beyond $t = 25$ is below 9×10^{-6}, and the 355/113 and 22/7 curves agree within 2×10^{-3}. The three rat-growth estimators of Basu and Chib cited in Section 6.3 differ from the analytic value printed beside them by −0.0011, +0.0006, +0.0210. The total-variation distances from the cluster-count table of Frühwirth-Schnatter and Malsiner-Walli quoted in the same section are 0.454 for the two-cell binning and 0.423 for the three-cell binning.

For the $N = 7{,}931$ tuberculosis stratum we computed the evidence of the finite mixture model for a single categorical variable, with the $k = 8$ sign patterns read as that variable, up to $g = 3$ with $\ln Z$ error bounds below 6.2×10^{-9}. The two-class conditionally independent latent-class model on the same stratum requires a lattice of 1.32×10^{14} states, which would need 1.06 petabytes at eight bytes per state.

S4 The proportional regime: many states instead of many components

The closed forms of the main paper hold at every k and N jointly, so they also apply in a regime that fixed-model asymptotics cannot describe, in which the number of states k grows in proportion to the sample size N. We compute the evidence density in the proportional regime $N = \nu k$ at $g = 2$, measure the deficit of the log evidence below the maximized log likelihood along the crossover family $k \sim N^{\alpha}$, and show that this deficit saturates past $\alpha = 1$. These results are summarized in Section 4.5 of the main paper.

S4.1 The evidence density at $N = \nu k$

Let the variable have k states and take $N = \nu k$ observations at fixed ν, with $g = 2$. Proportional regimes of this kind are the classical setting of the statistical mechanics of learning, where free energies per degree of freedom are computed by the replica method as quenched averages over data (Seung et al., 1992), including for Bayesian clustering of mixture data (Mozeika and Coolen, 2020). For the one-component model the corresponding quantity is the large-alphabet redundancy of universal coding, whose asymptotics in this regime are known with explicit constants (Orlitsky et al., 2004; Szpankowski and Weinberger, 2012). For the mixture, the singular case, we know of no prior computation. Because the evidence is available exactly, the large-k limit can be computed from one deterministic count vector at each k, with no disorder average and no replica calculation, and at every finite k the evidence is a rational number.

Define the evidence density

$$f(\nu) = \lim_{k \to \infty} \frac{\ln Z(U_k) + N \ln k}{k}, \qquad N = \nu k, \tag{S4.1}$$

on deterministic count rays $U_k = N q_k$ of fixed shape. The term $N \ln k$ is minus the log likelihood of the uniform distribution on the k states, so f is the limiting log evidence per state measured from that baseline. Because the one-variable mixture model contains every distribution on the k states, the maximized likelihood on the ray is exactly $\prod_v q_v^{N q_v}$, which caps the density at $f(\nu) \le \nu\,\mathrm{KL}$, with KL the divergence of the shape from uniform. The deficit $D(\nu) = \nu\,\mathrm{KL} - f(\nu)$ is the amount, per state and in the large-k limit, by which the log evidence falls below the maximized log likelihood. Its growth with ν is controlled by the learning coefficient of the singular model.

We measured f on two shapes. The two-rate shape has $q_v = \frac{3}{2k}$ on half the states and $\frac{1}{2k}$ on the other half, an interior truth. The two-block shape has $q_v = \frac{2}{k}$ on half the states and zero on the other half, a truth on a boundary stratum whose codimension grows with k. We evaluated the evidence at twelve values of ν from 2 to 200 on sequences of k up to 1600. Extrapolation in k with fitted $1/k$ and $(\ln k)/k$ corrections gives f to better than 2×10^{-4} at every ν, and to a few parts in 10^5 for $\nu \geq 3$. Over this window the deficit grows logarithmically, $D(\nu) = c \ln \nu + d + e/\nu$, with

$$c_{\text{two-block}} \;=\; 0.7498 \pm 0.0027\,, \qquad c_{\text{two-rate}} \;=\; 0.506 \pm 0.018\,, \tag{S4.2}$$

the errors dominated by the spread across fit forms.

The coefficient c is the limiting density of the learning coefficient of singular learning theory. Indeed, when the fixed-k expansion of Watanabe (Section 2.2 of the main paper) is inserted into Eq. (S4.1), the term $\lambda(k)\ln N$ contributes $(\lambda(k)/k)(\ln \nu + \ln k)$; the $\ln k$ part joins the corrections removed by the k-extrapolation, and the $\ln \nu$ coefficient of the deficit is the limiting density $\lim_k \lambda(k)/k$. At an interior truth the learning coefficients of Watanabe and Watanabe (2022) for two-component multinomial mixtures give $\lambda(k)/k \to 1/2$. The two-block truth violates their interior hypothesis, and the boundary-corrected density at Dirichlet concentration β is $1/4 + \beta/2$, hence $3/4$ at the flat priors used here. This corrected density is the coefficient of k in the boundary-strata formula for $\lambda(k)$ given in Section S4.2, and we checked it against a saturated model whose evidence has a closed form. Each measured coefficient agrees with the density predicted for its shape: $c_{\text{two-block}}$ lies 0.06 standard errors from $3/4$, with the uncorrected value $1/2$ far outside its error bar, and $c_{\text{two-rate}}$ is consistent with $1/2$ (0.35 standard errors). The prior dependence is itself a prediction: at Jeffreys priors ($\beta = 1/2$) the two-block coefficient should drop to $1/2$ while the two-rate one stays. The test at $k = 8$ is reported in Section S4.2: the flat-minus-Jeffreys difference of $\lambda(8)$ is 1.998 ± 0.004 against the predicted 2 for the two-block shape and 0.00000 ± 0.00002 against 0 for the two-rate shape. When the count shape is interpolated between the two-rate and the two-block form, the deficit varies smoothly, with no sign of a transition at the resolution of these measurements.

S4.2 Scaling the states with the sample: $k \sim N^{\alpha}$

The fixed-k expansion and the proportional regime are the two ends of a one-parameter family of regimes, the deterministic count rays with $k \sim N^{\alpha}$, $\alpha \in [0, 1]$. For $\alpha < 1$ the density f has no finite limit, so we work with the total deficit, $\Delta(N, k) = N\,\mathrm{KL} - (\ln Z + N \ln k)$, which equals $k\,D(\nu)$ at $N = \nu k$. Wherever fixed-k theory applies, $\Delta = \lambda(k)\ln N + \mathcal{O}(1)$, and we write $\bar\lambda(\alpha)$ for the measured coefficient of $k \ln N$ along a ray. We measured the map from α to $\bar\lambda(\alpha)$ on the two shapes, on rays $\alpha = 1/3, 1/2, 2/3$ (sequences to $k = 1600$, $N = 262{,}144$, every value of Z an exact rational), with the $\alpha = 0$ end refit at fixed $k \in \{4, 8, 16, 32\}$ and the $\alpha = 1$ end given by Eq. (S4.2).

A natural null hypothesis is a flat map, $\bar\lambda(\alpha) = 3/4$ (two-block; $1/2$ two-rate) independent of α, which matches both known ends. On the two-block shape the measured $\bar\lambda$ instead falls linearly with α: from 0.7567 ± 0.0034 at the fixed-k endpoint ($\lambda(32)/32$) through 0.5018 ± 0.024, 0.3751 ± 0.0035, and 0.2484 ± 0.0024 at $\alpha = 1/3$, $1/2$, $2/3$, against $\frac{3}{4}(1 - \alpha) = 0.500$, 0.375, 0.250; the flat map is excluded by up to 207 standard errors. The measurements follow a linear crossover,

$$\bar\lambda(\alpha) \;=\; c\,(1 - \alpha)\,, \tag{S4.3}$$

with c the coefficient of Eq. (S4.2) and no transition at any interior α. Equation (S4.3) is an empirical relation measured at three interior exponents. It appears to follow from an identity and one measured fact: along a ray, $k \ln \nu = (1 - \alpha)\,k \ln N$ exactly, so Eq. (S4.3) is equivalent to the statement that the deficit has the same coefficient c of $k \ln \nu$ at every α. The fit with $k \ln \nu$ as regressor confirms this: the coefficient is 0.7528 ± 0.0359, 0.7502 ± 0.0069, 0.7452 ± 0.0073 at $\alpha = 1/3$, $1/2$, $2/3$, each within 0.6 standard errors of the value 0.7498 ± 0.0027 in Eq. (S4.2). The deficit per state thus seems to depend on N and k only through their ratio $\nu = N/k$: every regime with $0 < \alpha < 1$ follows the $\alpha = 1$ law of Section S4.1 rather than fixed-k growth in $\ln N$.

At the fixed-k end ($\alpha = 0$) the two-block learning coefficient $\lambda(k)$ can be determined directly, including its k-independent term, which is divided by k in Eq. (S4.1) and is therefore absent from $f(\nu)$. The measured sequences give $\lambda(k) \to \frac{3}{4}k + \frac{1}{2}$ (the largest-k local slopes minus $\frac{3}{4}k$ are $+0.509$, $+0.498$, $+0.492$, $+0.481$ at $k = 4, 8, 16, 32$), which fixes at $+\frac{1}{2}$ the $\mathcal{O}(1)$ intercept that the boundary-corrected density left open. The line itself can be derived from the boundary strata of the two-block truth: $\lambda(k) = (\beta/2 + 1/4)\,k + \alpha_0 - \frac{1}{2}$ at component concentration β and mixing-weight concentration α_0, hence $\frac{3}{4}k + \frac{1}{2}$ at the flat priors used here. The intercept depends on the priors alone: it is the contribution of the mixing-weight prior at the dominant boundary stratum, where one mixing weight vanishes, less the contribution of the one Gaussian direction that the boundary face removes. In exact small-N enumeration the same intercept appears on deterministic and on sampled sequences. Its dependence on α_0 gives a falsifiable prediction: at a Jeffreys mixing prior ($\alpha_0 = \frac{1}{2}$) the intercept is 0. At $k = 8$ we tested both this prediction and the β-dependence of the line, using a local-slope extrapolation whose accuracy on the flat-prior sequences is 10^{-5} (it recovers the flat-prior values 6.5 and 3.5 of the two shapes). Under a Jeffreys component prior ($\beta = \frac{1}{2}$) at flat mixing, the two-block $\lambda(8)$ is 4.5004 ± 0.0007 against the line's 4.5 and the two-rate value is 3.500000 ± 0.000002 against 3.5. With the mixing prior also set to Jeffreys ($\alpha_0 = \frac{1}{2}$, component prior kept at $\beta = \frac{1}{2}$) the two-block value is 4.0000 ± 0.0001 against 4.0 (intercept 0), and the shift from $\alpha_0 = 1$ is 0.5004 ± 0.0008 against $\frac{1}{2}$, so each of these values at $k = 8$ agrees with its prediction.

On the interior (two-rate) shape the measured $\bar{\lambda}(\alpha)$ is 0.3315 ± 0.0079, 0.2500 ± 0.0001, 0.1648 ± 0.0017 at $\alpha = 1/3$, $1/2$, $2/3$, i.e. the linear law Eq. (S4.3) at $c = 1/2$ (the flat map excluded by up to 2593 standard errors). When the deficit is refitted with $k \ln \nu$ in place of $k \ln N$, its coefficient is again independent of α. At the fixed-k end of this shape the measured coefficient is $\lambda(k) = (k-1)/2$: 1.4718 ± 0.0107, 3.4271 ± 0.0276, 7.3357 ± 0.0621, 15.1516 ± 0.1321 at $k = 4, 8, 16, 32$, with largest-k local slopes 1.4981, 3.4950, 7.4887, 15.4762, and at $k = 2$ a slope of 0.4996 against a reference value of 3/4. The reference values, $k/2$ for $k \geq 3$ and 3/4 at $k = 2$, are those given by the theorem of Watanabe and Watanabe (2022) for mixtures of multinomials with $M \geq 2$ trials per observation, applied here to an $M = 1$ model. The discrepancy is exactly their component-splitting cost $\min(\alpha_0/2, (k-1)/4)$, $+\frac{1}{2}$ at every $k \geq 3$ and $+\frac{1}{4}$ at $k = 2$ at flat priors. That cost is attached to parameter directions along which the one-draw-per-observation likelihood considered here is constant, so it vanishes, and $\lambda(k) = (k-1)/2$ on deterministic rays and sampled data alike. A check with sampled data at $k = 2$ gives $\lambda = 0.4848 \pm 0.0063$, far from the reference 3/4 (41.9 standard errors) and 2.4 standard errors below 1/2, with the same finite-N lag as the deterministic sequences. The difference between deterministic and sampled data lies entirely in the $\mathcal{O}(1)$ term, flat in N at the derived value $-\frac{1}{2}$. The splitting cost $+\frac{1}{2}$ and the two-block intercept $+\frac{1}{2}$ are different quantities that coincide only at flat priors. At the Jeffreys mixing prior the intercept moves to 0 whereas the splitting cost would be $\frac{1}{4}$, and the $k = 8$ measurement above, 4.0000 ± 0.0001, agrees with the vanishing intercept and excludes the value 4.25 that a splitting cost of $\frac{1}{4}$ would give.

Past $\alpha = 1$, where the states outnumber the observations, the deficit saturates at the scale $N\,\mathrm{KL}$ set by the maximized likelihood. As $k \to \infty$ at fixed data support, $\ln Z + N \ln k$ converges to the finite, exactly computable constant $\ln Z_\infty(U) = \ln\big[\prod_v u_v!\, \sum_m R_U(m)\, m!\,(N-m)!/(N+1)!\big]$. Hence $\tilde{\Lambda} = \Delta/(N\,\mathrm{KL}) \to 1$: the deficit asymptotically equals the whole of $N\,\mathrm{KL}$, the information needed to localize the occupied states, and the evidence density per state tends to zero. We computed $\tilde{\Lambda}$ on sparse sequences ($N = 64$ to 512, $k = 4N$ to $256N$) of two shapes, with count

2 on $N/2$ states or count 1 on N states. On every sequence of either shape the computed values of $\ln Z + N \ln k$ converge to $\ln Z_\infty(U)$ (for the count-2 shape $\ln Z_\infty = 172.597657$ at $N = 512$). For the count-2 shape $\tilde{\Lambda}$ rises toward its ceiling ($0.8932 \to 0.9463$ at $N = 512$), and the finite-k residual of $\ln Z + N \ln k$ approaches zero as $-N^2/2k$: the largest-k residuals divided by N^2/k are -0.4729, -0.4870, -0.4933, -0.4964 across its four sequences. For the count-1 (singleton) shape $Z_\infty = 1$ exactly at every N (the identity $\sum_m \binom{N}{m} m!(N-m)! = (N+1)!$), the ceiling is approached from above ($1.0457 \to 1.0002$), and the coefficient of N^2/k in its residual is -0.327 to -0.329, near $-1/3$; we know of no reference value for it.

S5 Capture–recapture: population size and number of classes

Ecologists use closed-population capture–recapture to estimate the size of an animal population that cannot be counted directly. Animals are trapped, marked, released, and trapped again over T occasions; the data are the capture histories of the D animals seen at least once; the quantity of interest is the total number of animals N, including those never seen. (Throughout this section N denotes the population size, as is conventional in capture–recapture, and not a sample size as in the main paper.) When every animal is equally catchable, estimation of N under the resulting homogeneous model is classical. Real animals are not equally catchable, and the standard remedy, the finite-mixture model M_h of Pledger (2000), splits the population into g latent classes with class-specific capture probabilities. Link (2003) then proved that heterogeneity models fitting the same data essentially equally well can imply very different N, so the estimate of N depends on the assumed number of classes, which the data do not identify. Joint Bayesian inference on N and g has so far been carried out by reversible-jump MCMC, which Arnold et al. (2010) applied to the same two classic datasets used below; such a sampler approximates the posterior that is computed exactly in this section.

We compute the joint posterior $P(N, g \mid \text{data})$ for $g = 1, 2, 3$ latent classes on two datasets with $T = 6$ and $T = 18$ occasions and 68 and 76 animals seen (Table 6 of the main paper). The model is the tied-p finite mixture: N animals, latent class $a \in \{1, \ldots, g\}$ with weights $w \sim \text{Dir}(\gamma_0)$, class capture probability $p_a \sim \text{Beta}(b_1, b_2)$ shared across the T occasions, per-animal capture count binomial, and zero-count animals unobserved. The prior on N is uniform or $1/N$ on $\{1, \ldots, \text{cap}\}$ with cap $= 300$ or 400; because truncation at the cap removes any posterior mass beyond it, we give both caps side by side. The hyperprior takes four settings, $H_1 = (\gamma_0{=}1, b_1{=}1, b_2{=}1)$, $H_2 = (1, 2, \frac{1}{2})$, $H_3 = (2, 1, 1)$, and $H_4 = (2, 2, \frac{1}{2})$ in the same order, which, crossed with the two N-priors and the two caps, gives sixteen prior settings per dataset and g. Posterior probabilities of g use a uniform prior on $g \in \{1, 2, 3\}$. The one- and two-class posteriors are computed in exact rational arithmetic, and the three-class posteriors in floating point by sums that involve no subtraction and are therefore free of cancellation error. Against exact rational values this floating-point evaluation has relative error at most 7.1×10^{-14} over 912 comparisons on small tables, 1.1×10^{-13} on complete three-class posteriors for reduced hare tables (20 animals at all four hyperpriors and 30 animals at H_1), and 8.1×10^{-13} on the full-scale two-class hare and rabbit posteriors (325 to 333 population values each, all four hyperpriors). The full-scale three-class posteriors were not checked directly against an exact computation. A three-class evaluation costs roughly 10^{11} to 4×10^{11} arithmetic operations per dataset.

Both datasets analyzed here are standard examples in the capture–recapture literature. The snowshoe hare data have $T = 6$ occasions, $D = 68$ animals, $S = 145$ captures, and frequency vector $f = (25, 22, 13, 5, 1, 2)$ (Baillargeon and Rivest, 2007). The Edwards–Eberhardt cottontail rabbit data have $T = 18$ occasions, $D = 76$ animals, $S = 142$ captures, and frequency vector $f = (43, 16, 8, 6, 0, 2, 1)$ with no animal caught more than seven times, from a penned population whose true size, $N = 135$, is known (Edwards and Eberhardt, 1967). The posterior of N under each number of classes can therefore be compared with the true value, a check rarely available in a model-selection problem. (A later restatement of these data lists 75 animals (Matechou et al., 2016); the data as distributed and the printed classical estimates computed from them both give $D = 76$.)

g	cap	mode N	$E[N]$	sd	central 95%	$P(N \leq 135)$
1	300	96	97.8894	7.203	$[86, 114]$	0.99989703
1	400	96	97.8894	7.203	$[86, 114]$	0.99989703
2	300	118	147.3991	41.834	$[98, 263]$	0.49977568
2	400	118	153.8321	54.188	$[98, 317]$	0.48334254
3	300	120	149.4173	41.777	$[99, 264]$	0.47181353
3	400	120	155.2640	52.985	$[99, 312]$	0.45747743

Table S1: Posterior of the population size N for the rabbit data (known $N = 135$) under $g = 1, 2, 3$ latent classes, at the flat reference prior (H_1, uniform N-prior) and both truncation caps. With one class the known size lies above the 99.98th percentile of the posterior; with two or three classes it lies near the middle of the central 95% interval.

Table S1 compares the rabbit posteriors with the known population size. Under the homogeneous $g = 1$ model the true size lies above the 99.98th percentile of the posterior, and outside the central 95% interval, under every one of the sixteen prior settings. The failure of the homogeneous analysis on these data therefore does not depend on which of these priors is used. Under $g = 2$ the true size lies between the 45th and 79th percentile across the prior settings and inside every central 95% interval; under $g = 3$ it lies between the 44th and 82nd percentile, again inside every interval. The marginal likelihoods lead to the same conclusion as the percentiles: they favor heterogeneity under every prior setting, with $\log_{10} \mathrm{BF}(2{:}1)$ between 0.546 and 1.988, Bayes factors of roughly 3.5 to 97 (Table S2). A third class is never favored: $\log_{10} \mathrm{BF}(3{:}2)$ is negative under every rabbit prior setting, between -1.580 and -0.297. The homogeneous maximum-likelihood estimate for the rabbits is 96 against a true size of 135, a standard illustration of heterogeneity bias; here the whole one-class posterior of N shows the same bias.

The hare data have no known population size, and for them the number of latent classes is less well determined than for the rabbits. With the uniform prior on g, the homogeneous model retains at most 0.12 posterior probability under every prior setting, so the posterior puts at least 0.88 on two or three classes throughout. However, whether two or three classes are favored depends on the hyperprior: $\log_{10} \mathrm{BF}(3{:}2)$ runs from -0.291 to $+0.216$ across the prior settings. At $T = 6$ occasions the three-class model is only just identified, and the data could not be expected to distinguish two classes from three. The posterior mean of N under the flat reference prior rises with the number of classes, from 75.78 (g=1) through 84.20 (g=2) to 89.54 (g=3) at cap 400, with 75.78, 83.61, 88.84 at cap 300. This rise is a finite-sample measure of the sensitivity of N to the assumed number of classes; Link derived that sensitivity from an asymptotic argument.

Exact Bayesian posteriors are classical for the homogeneous models: Castledine (1981) and Smith (1991) gave the N-posteriors for the constant-probability model M_0 and the time-varying

dataset	hyp	N-prior	$\log_{10}$ BF(2:1)		$\log_{10}$ BF(3:2)	
			cap 300	cap 400	cap 300	cap 400
hare	H_1	uniform	1.061254	1.062245	0.215580	0.215781
hare	H_1	$1/N$	1.027357	1.027594	0.195036	0.195100
hare	H_2	uniform	1.115027	1.115034	−0.220172	−0.220160
hare	H_2	$1/N$	1.099051	1.099053	−0.228237	−0.228234
hare	H_3	uniform	0.915892	0.917087	0.197629	0.196954
hare	H_3	$1/N$	0.868318	0.868614	0.182232	0.182071
hare	H_4	uniform	0.710183	0.710197	−0.283258	−0.283265
hare	H_4	$1/N$	0.688993	0.688997	−0.290957	−0.290958
rabbit	H_1	uniform	1.973079	1.987599	−0.297478	−0.298597
rabbit	H_1	$1/N$	1.821508	1.827449	−0.303820	−0.304183
rabbit	H_2	uniform	0.837960	0.840035	−1.070112	−1.071031
rabbit	H_2	$1/N$	0.734826	0.735583	−1.063256	−1.063597
rabbit	H_3	uniform	1.874070	1.886116	−0.381663	−0.384502
rabbit	H_3	$1/N$	1.716853	1.721859	−0.385179	−0.386312
rabbit	H_4	uniform	0.659781	0.661738	−1.578154	−1.579406
rabbit	H_4	$1/N$	0.546434	0.547167	−1.567425	−1.567899

Table S2: $\log_{10}$ Bayes factors between the latent-class models on both datasets under all sixteen prior settings (hyperprior, N-prior, and truncation cap 300 or 400). Both datasets reject homogeneity under every setting; no rabbit setting favors a third class; for the hares the sign of BF(3:2) changes with the hyperprior. The cap changes these values by at most 0.015.

model M_t with Beta priors, model-averaged N-posteriors followed for log-linear models of list dependence (Madigan and York, 1997; King and Brooks, 2001), and a 2026 result gives the two-sample hypergeometric posterior, again with no heterogeneity (Mirzaee et al., 2026). The Bayesian latent-class treatment of multi-list capture, the design of the die study in Section S6, has its own sampler-based literature, from the multiple-lists population-size analysis of Fienberg et al. (1999) to the Dirichlet-process latent-class estimator of Manrique-Vallier (2016). The exact marginal-likelihood kernel for fixed-N mixtures of independence models is due to Lin et al. (2009), and exact weight-posteriors for known-component mixtures appeared in 2026 (Meshcheryakov, 2026; Kim, 2026). To our knowledge, the joint posterior $P(N, g \mid \text{data})$ and its cross-model Bayes factors for these latent-class models, with the unknown N and the unobserved zero-count cell inside the integral, have not previously been computed exactly.

S6 Counting the dies of a Roman mint

Ancient coins were struck by hand between engraved dies, and how many dies a mint used is a question numismatists have asked for a century. Die counts are used to estimate how much money an ancient state produced, and whether they can support such estimates has been disputed (Buttrey, 1993; de Callataÿ, 1995). The denarii of P. Crepusius, a Roman Republican issue, are a standard test case because the mint numbered its own reverse dies, engraving a control numeral on each, and in his die study Buttrey read numerals up to 519 (Buttrey, 1976). Because the number of dies is known from these numerals, the issue provides a second test of the latent-class capture models of Section S5 against a known population size, in addition to the Edwards–Eberhardt rabbit data analyzed there. The estimator numismatists use in practice is Esty's frequentist coverage model (Esty, 2006, 2011), applied today by the Roman Republican Die Project to this same issue (Carbone and Yarrow, 2020). Statisticians have surveyed such estimators (Stam, 1987) and, for the related problem of estimating the number of species, have derived objective-Bayes posteriors for mixed-Poisson abundance data (Barger and Bunge, 2010).

The data are the open linked-data records of Schaefer's die attributions, published by the American Numismatic Society and nomisma.org under the Open Database License. They contain 3,304 die-attributed specimens of the issue (one specimen excluded for bearing two reverse-die attributions), with $D = 413$ distinct reverse dies observed. In the natural capture design the hoards in which specimens of the issue were found would be the capture occasions, with the data arranged as a die-by-hoard table built from per-specimen hoard links. That table does not exist in the open data: only 34 of the 3,304 die-attributed specimens carry an open per-specimen hoard link (1.03%, six hoards). The catalog's figure of 82 hoards is a count of the hoards known to contain the coin type, given without specimen-level links. The open data do support a multi-list structure. Fifteen modern holding datasets serve as the capture lists, a die being captured by a list if at least one of its specimens is held there; this gives $S = 829$ list-captures and capture-frequency vector $f = (150, 145, 87, 28, 2, 1)$ indexed by the number of lists, from one to six, that capture the die. We therefore compute $P(N, g \mid f)$ from the capture-frequency histogram of the fifteen holdings under the tied-p exchangeable-list mixture of Section S5, with $g = 1$ and 2 latent classes (Table 6 of the main paper). As before, N is the unknown being estimated, here the number of dies including those never seen. One holding contains 2,755 of the 3,304 specimens and 95.4% of the observed dies, so capture effort is far from equal across the lists, contrary to the model's assumption, and the results below should be read with that limitation in mind.

The true number of dies comes from the die records themselves and is not used in the computation. Each die record carries Schaefer's reading of the die's control numeral. The highest attested numeral is 519, with 405 distinct numerals attested and 114 missing, so the number of numerals issued is 519 and the number of dies is at least 519 (a few dies share a numeral under letter suffixes). The computation uses the floating-point routine of Section S5; on a fifteen-list subtable of the same shape ($D = 40$ dies, $g = 1$ and 2, both hyperprior settings) the routine agrees with exact rational evaluation to relative error 7.0×10^{-13}. Only $g = 1$ and 2 are computed for the die data: at the size of the full die table the three-class state space is more than a million times larger than the two-class one.

At $g = 1$ the posterior has mode 497 under every prior setting of Table S3, with central 95% set $[475, 526]$ under six settings and $[475, 525]$ under the two $1/N$ settings at P_2. The posterior reaches the attested count of 519 only at its upper edge, with $P(N \geq 519) = 0.07$, which places 519 at the 93rd percentile. At $g = 2$ the posterior widens upward and 519 lies inside the central 95% set under every setting, with $P(N \geq 519)$ between 0.10 and 0.21. However, the evidence computed from the capture-frequency histogram favors one class: $\log_{10} \mathrm{BF}(2{:}1)$ runs from -0.99 to -2.21 across the prior settings, and the posterior probability of two classes is below 0.1 under every setting. Unequal capture effort across the lists cannot produce a spurious two-class signal, but it can mask genuine heterogeneity among the dies. The preference for one class may therefore reflect the unequal sizes of the fifteen holdings more than homogeneous output across the dies of the Crepusius mint.

On the same issue-level counts (all 3,304 specimens and 413 dies of the issue), Esty's estimator gives 472 in its simplest form and 478 with 95% interval $[469.4, 487.3]$ in the form the die-study literature applies. The public implementation in current use, run on the largest subtype's reverses,

cap	hyp	N-prior	g	mode	$E[N]$	central 95%	$P(N \geq 519)$	$\log_{10}$ BF(2:1)
1000	P_1	uniform	1	497	499	$[475, 526]$	0.074	
1000	P_1	uniform	2	500	513	$[477, 661]$	0.199	-1.00
1000	P_1	$1/N$	1	497	499	$[475, 526]$	0.071	
1000	P_1	$1/N$	2	499	510	$[477, 611]$	0.180	-1.01
1000	P_2	uniform	1	497	499	$[475, 526]$	0.070	
1000	P_2	uniform	2	499	502	$[476, 533]$	0.116	-2.21
1000	P_2	$1/N$	1	497	498	$[475, 525]$	0.067	
1000	P_2	$1/N$	2	498	501	$[476, 532]$	0.109	-2.21
1500	P_1	uniform	1	497	499	$[475, 526]$	0.074	
1500	P_1	uniform	2	500	520	$[477, 732]$	0.206	-0.99
1500	P_1	$1/N$	1	497	499	$[475, 526]$	0.071	
1500	P_1	$1/N$	2	499	513	$[477, 631]$	0.183	-1.01
1500	P_2	uniform	1	497	499	$[475, 526]$	0.070	
1500	P_2	uniform	2	499	502	$[476, 533]$	0.116	-2.21
1500	P_2	$1/N$	1	497	498	$[475, 525]$	0.067	
1500	P_2	$1/N$	2	498	501	$[476, 532]$	0.109	-2.21

Table S3: Posterior of the number of dies N (the mint's own count is 519) under one and two latent classes, for truncation caps 1000 and 1500, hyperprior settings $P_1 = H_1$ and $P_2 = H_4$ of Section S5, and both N-priors, with the $\log_{10}$ Bayes factor of two classes against one. $P(N = \text{cap})$ is below 10^{-4} under every setting and below 10^{-68} at $g = 1$.

gives 462 with interval $[452, 472]$, so every issue-level application of Esty's estimator excludes the attested 519. Esty's estimator is a function of the specimens-per-die abundance histogram, which summarizes the same specimens differently from the list-capture table used here, so the two analyses can be compared only through the die totals each supports. Run on the small subtype 1a alone, the same implementation gives 519 with interval $[388, 695]$; that figure, however, is an estimate of the number of dies in the subtype only and is not a second issue-level count. The Roman Republican Die Project has published its own comparison of the same estimator with the numbered dies. In that comparison Esty's interval implies 48 to 65 missing dies, whereas the project counts 115 missing from the numeral sequence (one more than the 114 in the open die records used here) and calls the predicted range "way too low" (Carbone and Yarrow, 2020). By contrast, the central 95% set of the multi-list posterior contains 519 under every prior setting considered; to our knowledge no Bayesian or exact-posterior die count has previously appeared in the numismatic literature.

We also analyzed the partial die-by-hoard table that can be assembled from the open hoard links ($T = 6$ hoards, $D = 32$ dies, capture-frequency vector $(31, 1, 0, 0, 0, 0)$). With 31 of the 32 dies appearing in only one hoard each, the posterior follows the prior and the cap rather than the data. Across the same sixteen prior settings at caps 700 and 1200 its mode runs from 98 to 229 and its mean from 146 to 490, so the open hoard table does not constrain N. A die count is not by itself an estimate of coin production: converting dies to coins struck requires a coins-per-die multiplier, which is the quantity disputed in the production controversy (Buttrey, 1993; de Callataÿ, 1995) and is outside the scope of this analysis.

S7 Latent classes in an environmental-DNA panel

Environmental DNA surveys detect rare species by amplifying their DNA from filtered water instead of trapping the animal. The data form replicate panels: each water sample is run through T PCR replicates, and each replicate either amplifies the target DNA or does not. Such panels are usually analyzed with occupancy models that give each sampling unit two latent states, target DNA present or absent, together with detection and false-positive error probabilities (Lahoz-Monfort et al., 2016). How many latent classes the panel itself supports is a model-selection question. We address it on a standard tidewater-goby panel of $N = 356$ samples whose detection-count histogram has $k = 7$ cells (Table 6 of the main paper). We compute the evidence for $g = 1, 2, 3$ latent sample classes, the posterior $P(g \mid \text{panel})$, and the evidence of a continuous alternative in which the detection probability varies from sample to sample.

The panel is the tidewater-goby detection table distributed with the eDNAoccupancy package (Dorazio and Erickson, 2018): 356 water samples scored in $T = 6$ PCR replicates each, across 39 site labels, from the regional survey of this endangered species by Schmelzle and Kinziger (2016), whose own occupancy analysis used 29 locations. The detection-count histogram, from zero to six amplifications out of six, is $f = (197, 14, 16, 10, 17, 19, 83)$, 737 amplifications in all; the six replicate totals, 120, 124, 120, 127, 125 and 121, are nearly equal, consistent with the exchangeable-replicate assumption under which the histogram summarizes the panel. The model is the exchangeable-sample latent-class mixture: each sample belongs to one of g classes with weights $w \sim \text{Dir}(\gamma_0)$, and a sample in class a amplifies each replicate independently with probability $\theta_a \sim \text{Beta}(\beta, \beta)$. In contrast with the capture–recapture setting of Section S5, the number of samples is known here, so the evidence integral contains no unobserved zero cell and only the number of classes remains to be inferred. The priors considered are $\beta \in \{1, \frac{1}{2}\}$ crossed with $\gamma_0 \in \{1, \frac{1}{2}\}$, four settings, and posterior model probabilities use equal prior weights on the models compared. The one- and two-class evidences are computed in exact rational arithmetic and the three-class evidences in floating point. The floating-point routine reproduces the exact two-class evidence on this table, both directly and as a three-class computation with one class emptied, to relative error below 4.6×10^{-13}, and a floating-point evaluation of the closed-form beta-binomial evidence agrees with an exact-rational one to 1.1×10^{-15}.

The one-class fit puts its mass in the middle of the histogram, predicting 28.2 all-negative samples where 197 are observed and 0.6 all-positive samples where 83 are observed (Table S4). The one-class model is rejected outright, with $\log_{10} \text{BF}(2{:}1)$ between 292.50 and 293.11 across the four prior settings (Table S5). Two classes are also clearly insufficient, with $\log_{10} \text{BF}(3{:}2) = 15.5$ to 16.1 under all four settings. The two-class fit reproduces most of the counts in the two large end cells but underestimates the middle ones, predicting 3.8 and 2.7 samples in the two- and three-amplification cells where 16 and 10 are observed (Table S4). A four-class model is not identifiable at $T = 6$ ($2g - 1 = 7$ parameters against the six free frequencies of the seven-cell histogram), which limits the comparison to three classes or fewer, so among finite mixtures the panel requires at least three latent sample classes. With equal prior weights on $g = 1, 2, 3$ the posterior probability of $g = 3$ exceeds $1 - 10^{-15}$ under every setting.

The three-class model has the highest evidence among the finite mixtures but a lower evidence than the continuous alternative, per-sample beta-binomial mixing with hyperparameters (a, b) given a uniform prior on the 7×7 grid $\{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8\}^2$. The continuous model is preferred under every prior setting, with $\log_{10} \text{BF}(3{:}\text{BB})$ from -1.98 to -2.96 against the hierarchical evidence and from -3.67 to -4.65 against the grid maximum (Table S5). With equal prior weights on the four

amplifications	0	1	2	3	4	5	6
observed	197	14	16	10	17	19	83
one class	28.2	88.8	116.8	82.2	32.6	6.9	0.6
two classes	183.5	39.1	3.8	2.7	15.0	47.7	64.2
beta-binomial	186.8	26.7	17.7	15.4	16.0	21.2	72.3

Table S4: Observed detection-count histogram of the goby panel and posterior-predictive expected counts, $E[f_k] = 356\binom{6}{k}\, Z(f+e_k)/Z(f)$, under one class, two classes and the beta-binomial model, at the flat prior ($\beta = 1$, $\gamma_0 = 1$; the other prior settings move every entry by less than one sample). The one-class model fills the middle cells, which are nearly empty in the data; the beta-binomial tracks all seven cells.

β	γ_0	$\log_{10}$ BF(2:1)	$\log_{10}$ BF(3:2)	$\log_{10}$ BF(3:BB)	$\log_{10}$ BF(3:BB$_{\max}$)
1	1	+292.68	+15.65	−2.68	−4.37
1	$\frac{1}{2}$	+292.50	+15.55	−2.96	−4.65
$\frac{1}{2}$	1	+293.11	+16.10	−1.98	−3.67
$\frac{1}{2}$	$\frac{1}{2}$	+292.93	+15.99	−2.26	−3.95

Table S5: $\log_{10}$ Bayes factors on the goby panel under the four prior settings: two classes against one, three against two, and three classes against the beta-binomial model, scored by its hierarchical evidence over the 7×7 hyperparameter grid (BB) and by its grid maximum (BB$_{\max}$). Negative entries in the last two columns favor the beta-binomial (see text).

models, the beta-binomial carries between 98.9% and 99.9% of the posterior probability across the prior settings. The hierarchical evidence is dominated by a single point of the 49-point grid, the maximizer $(a, b) = (\frac{1}{8}, \frac{1}{4})$, so the hierarchical margin is smaller than the grid-maximum margin by $\log_{10} 49$. Since this maximizer lies on the edge of the grid, the grid maximum is only the most favorable of the hyperparameter values tried and should not be read as a bound over all continuous mixing densities. The evidences at the 49 grid points also show which mixing densities the panel supports. A uniform mixing density, $(a, b) = (1, 1)$, has $\ln Z = -871.2$ and the arcsine density $(\frac{1}{2}, \frac{1}{2})$ has -762.9, both far below the two-class evidences, which lie between -706.42 and -705.43. The Beta density at the maximizer $(\frac{1}{8}, \frac{1}{4})$ is U-shaped, with its mass near $\theta = 0$ and $\theta = 1$, the detection probabilities that produce the all-negative and all-positive samples dominating the histogram. Because the beta-binomial evidence has a closed form at each fixed (a, b), the average over the 49-point grid needs no numerical integration, whereas a continuous hyperprior on (a, b) would require a quadrature that we have not done. Continuous mixing has a precedent in eDNA analysis, where the DNA concentration underlying quantitative PCR data is modeled continuously (Jones et al., 2025).

All of the evidence comparisons above use the pooled detection-count histogram under the exchangeable-sample assumption. The 39-site grouping is not modeled, and site-level heterogeneity is a likely source of the sample-to-sample variation in detection probability that the beta-binomial model represents. The two-class model here is accordingly the exchangeable-sample one; the site-structured occupancy models of standard eDNA analysis are not among the models compared.

Latent detection classes in occupancy models have been studied before: Royle (2006) introduced finite mixtures of detection classes and chose among them with AIC, Royle and Link (2006) added a third, false-positive class, and Lahoz-Monfort et al. (2016) applied that model to eDNA replicate panels. All three fit by maximum likelihood and select among models with information criteria. Among Bayesian treatments, Sollmann et al. (2021) fitted a Dirichlet-process community occupancy model, which clusters species rather than samples and whose number of clusters was always overestimated in their own simulations. Martel et al. (2021) reanalyzed the same goby system with a Bayesian occupancy model whose two latent states were fixed a priori. A third latent state is therefore not new, and neither is a posterior over the number of latent detection classes: Turek et al. (2021) put a Dirichlet-process mixture on detection heterogeneity in occupancy and capture–recapture models and inferred the number of subgroups by MCMC. That posterior is sampler-based, however, and to our knowledge no exact evidence or exact posterior over the number of latent classes has been computed before for occupancy detection histories or eDNA replicate panels.

S8 Exact posteriors for Bayesian record linkage

Record linkage is the task of determining which of the noisy records in a file correspond to the same person, and hence how many distinct people the file refers to. In census returns, patient registries, and administrative rolls, for example, one person may appear several times with typos and transcription slips. The Bayesian formulation puts a prior on the partition of records into entities and a distortion model on the fields. Posterior inference under this formulation has relied on MCMC since the early Bayesian linkage model of Fortini et al. (2001): Metropolis-within-Gibbs (Tancredi and Liseo, 2011), Gibbs partitioning (Sadinle, 2014), the Gibbs sampler under empirically motivated priors of Steorts (2015), and the distributed partially collapsed Gibbs of dblink (Marchant et al., 2021). A standard benchmark is RLdata500: 500 German-style records with first name, last name, and birth date (year, month, and day) fields, 450 true entities, 50 duplicate pairs with generated errors. The published Bayesian posteriors for its number of entities come from MCMC samplers. Steorts (2015), with an empirically motivated prior, reports a posterior mean of 449 with standard deviation 7.2. The unified de-duplication framework of Tancredi et al. (2020) gives MCMC entity-count posteriors on RLdata500 and RLdata10000 under the same hit-and-miss likelihood family used below, with an extensive hyperparameter sensitivity analysis and an explicit proposal to investigate Pitman–Yor-type partition priors.

For RLdata500 ($N = 500$ records on five populated fields, 450 true entities; Table 6 of the main paper) we compute the posterior distribution of the number of entities under a fixed blocking of the records by birth year. We obtain per-block co-reference posteriors, the global entity-count posterior by convolution across all 86 blocks (Section 4.4 of the main paper), and Bayes factors between partition priors. The model is a hit-and-miss likelihood of Fellegi–Sunter class (Fellegi and Sunter, 1969; Copas and Hilton, 1990): a latent partition per block, entity field values drawn from the empirical field frequencies over all 500 records, and, on each of the five populated fields, a distortion under which a distorted draw may still hit the true value. The distortion probability is either fixed at a plug-in value ($\beta = 1/10$ per field) or integrated against a Beta prior ($\mathrm{Beta}(1, 4)$ per field, block-local, with a $\mathrm{Beta}(1, 9)$ sensitivity analysis). We compare four partition priors: uniform over block-refining partitions, Dirichlet process ($\alpha = 1$) (Ferguson, 1973), Pitman–Yor ($\theta = 1$, $\sigma = 1/2$) (Pitman and Yor, 1997), and the uniform-labelings prior of blink ($N_{\max} = 500$) (Steorts, 2015). Within each block the posterior quantities are computed by summing over all set partitions of its records; the largest block has 190,899,322 partitions, and across all blocks the enumeration totals 197,455,281 partition terms. The Bayes factors among these four priors form part of the RLdata500 column in the comparison of Section S9. On RLdata500 by itself the sign of these Bayes factors depends on the distortion model: the $\log_{10}$ Bayes factor for the Dirichlet process over Pitman–Yor is -11.23 under plug-in distortion and 261.86 under $\mathrm{Beta}(1, 4)$.

Blocking on birth year splits 8 of the 50 true duplicate pairs across blocks: even perfect within-block linkage yields 458 entities against the true 450, so 458 is the smallest count attainable under this blocking (the blocking floor), and posterior mass below 458 can only come from false merges. The published MCMC summary of 449 ± 7.2 (Steorts, 2015) was computed without blocking, so its distance from the blocked posteriors below reflects differences in blocking, partition prior, and likelihood rather than Monte Carlo error. The dblink authors warn that naive blocking “may induce significant error in the posterior” (Marchant et al., 2021); on this benchmark that error is the 8 entities between the floor and the truth. Under the finer (birth year, birth month) blocking, 371 blocks of at most four records, the floor rises to 462, with 12 of 50 pairs split.

distortion	prior	mean	sd	MAP	95% interval	$z(449)$
plug-in β=0.1	uniform	339.137	5.628	339	$[328, 350]$	19.52
plug-in β=0.1	DP	327.08	5.918	327	$[315, 339]$	20.6
plug-in β=0.1	PY	452.72	2.222	453	$[448, 457]$	-1.67
plug-in β=0.1	blink	438.284	3.182	438	$[432, 444]$	3.37
$\mathrm{Beta}(1, 4)$	uniform	198.849	5.795	199	$[188, 210]$	43.17
$\mathrm{Beta}(1, 4)$	DP	141.501	5.441	141	$[131, 152]$	56.52
$\mathrm{Beta}(1, 4)$	PY	442.217	3.877	443	$[434, 449]$	1.75
$\mathrm{Beta}(1, 4)$	blink	410.227	4.981	410	$[400, 420]$	7.78

Table S6: Posterior of the number of entities K on RLdata500 under birth-year blocking (truth 450, blocking floor 458), for four partition priors and two distortion models; the $\mathrm{Beta}(1, 9)$ sensitivity results are given in the text and in Figure S1. $z(449)$ is the distance of the published unblocked MCMC mean from each posterior, in posterior standard deviations.

Under birth-year blocking and plug-in distortion, the posterior mean number of entities on RLdata500 is 327.08 under the Dirichlet-process prior and 452.72 under Pitman–Yor, so at this sample size the partition prior appears to influence the posterior more than the data do. The Pitman–Yor posterior concentrates just below the floor (Table S6, Figure S1); the Dirichlet-process and uniform priors, whose partition weights favor merging at this block scale, pull the posterior hundreds of entities below both the floor and the published unblocked MCMC mean. With a $\mathrm{Beta}(1, 9)$ distortion prior (dashed in Figure S1) the posterior means are 266.0 (uniform), 219.6 (Dirichlet process), 452.6 (Pitman–Yor), and 435.0 (blink), in the same rank order as under plug-in and $\mathrm{Beta}(1, 4)$ distortion. Under the finer birth-year-and-month blocking we computed the same eight prior–distortion combinations: the means run from 393.8 to 460.1 under plug-in distortion and from 376.8 to 456.9 under $\mathrm{Beta}(1, 4)$, with Pitman–Yor again nearest the floor, now 462.

The per-pair co-reference posteriors, each the probability that two records in a block refer to one person, depend on the partition prior as strongly as the entity count does. Over the 1,545 within-block record pairs, all twelve prior–distortion combinations (four priors under plug-in, $\mathrm{Beta}(1, 4)$, and $\mathrm{Beta}(1, 9)$ distortion) rank the true pairs well (average precision 0.9457 to 0.9969 against within-block truth). At the 0.5 threshold every combination but two finds all 42 within-block true pairs (Pitman–Yor at plug-in and $\mathrm{Beta}(1, 9)$ distortion finds 41). However, the false-pair count at that

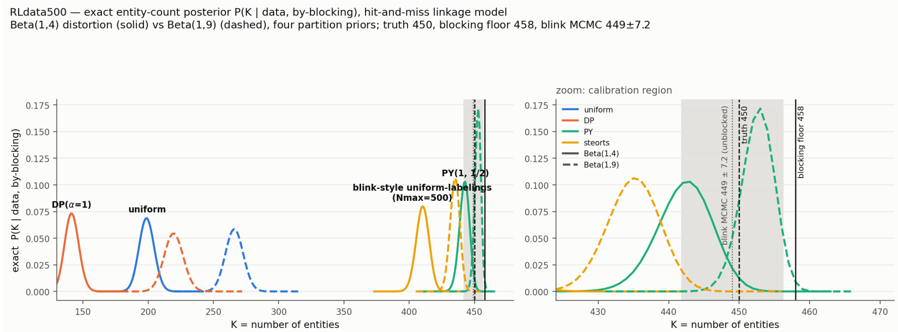

Figure S1: Entity-count posteriors $P(K \mid \text{data}, \text{blocking})$ on RLdata500 under birth-year blocking for four partition priors, with $\text{Beta}(1,4)$ distortion (solid) and $\text{Beta}(1,9)$ (dashed); vertical marks show the truth (450), the blocking floor (458), and the published unblocked MCMC summary (449 ± 7.2). The posteriors under the four priors differ by much more than their widths, and only the Pitman–Yor posterior has most of its mass near the floor.

threshold ranges from 4 (plug-in, Pitman–Yor) to 1,186 ($\text{Beta}(1,4)$, Dirichlet process), i.e. precision from 0.91 down to 0.034. The recall ceiling against all 50 true pairs is 0.84 throughout, since the 8 pairs split by the blocking are unreachable.

All posteriors in this section are conditional on the blocking, which excludes cross-block links, and on the five-field model. The plug-in computations use exact rational arithmetic at every block size. The Beta-integrated computations use exact rational arithmetic at blocks of up to eight records and, at the eighteen blocks of nine to fourteen records, double-precision enumeration with a quadrature rule exact for the degree of the integrand. The double-precision enumeration agrees with exact rational arithmetic to 5×10^{-14} at nine records, the largest block size where both were run, and assembling $P(K)$ in floating point reproduces the exact plug-in $P(K)$ to 3.4×10^{-13}. The $P(K)$ distributions are truncated at 10^{-12} per bin.

Linkage models used in practice include more structure than the five-field model of this section; for example, Steorts's model treats the name fields as strings under an edit-distance distortion, whereas ours treats every field as categorical (Steorts, 2015). The published Bayesian entity-resolution posteriors on RLdata500 and on other benchmarks were obtained by MCMC, and the survey of Binette and Steorts (2022) contains no instance of exact posterior computation. The exact computations we are aware of in this literature are conditional steps inside a sampler (the matching-matrix collapse of Tancredi and Liseo, 2011; within-block conditional enumeration as a Gibbs step, McVeigh et al., 2019), and full enumeration has been used to check partition samplers in clustering, though not in entity resolution (Bouchard-Côté et al., 2017). To our knowledge, the posteriors above are therefore the first computed exactly for a record-linkage model on a standard benchmark.

S9 An evidence comparison of partition priors

We rank eight partition priors by their Bayesian evidence on three record-linkage corpora. These are the RLdata500 benchmark of Section S8, its 10,000-record sibling RLdata10000, and the Italian household survey SHIW0810, with 39,743 records and 28,584 true entities (roughly 3×10^4 latent entities under the blocking used, since a true entity whose records fall in several blocks counts once in each block). The linkage likelihood uses five categorical fields for the two RLdata corpora and eight for SHIW0810. Every Bayes factor in this section is conditional on the blocking, on the hit-and-miss likelihood, and on the hyperparameter values, all specified below together with the sensitivity analysis around them.

Which partition prior a linkage model should use is a question with an active literature of its own. Duplicate detection needs clusters that stay small as the number of records N increases, whereas under the standard exchangeable priors, such as the Dirichlet process and Pitman–Yor, cluster sizes grow linearly in N. The microclustering priors introduced to meet this requirement are partition prior families under which cluster sizes grow sublinearly in N; the exchangeable sequences of clusters (ESC) priors are one such family (Miller et al., 2015; Zanella et al., 2016; Betancourt et al., 2022). To our knowledge none of this literature compares partition priors by Bayesian evidence on data. The priors are compared structurally and asymptotically (De Blasi et al., 2015), or by F1 against the true links after MCMC fitting (Marchant et al., 2023). The closest precedents are a sequential Monte Carlo approximation to the marginal likelihood, used to fit hyperparameters within one prior family (Di Benedetto et al., 2021), and Bayes factors between two candidate partitions under one fixed model, a theoretical device for impossibility results (Johndrow et al., 2018). In the model-selection literature, Casella et al. (2014) select partitions by Bayes factors under one fixed prior on partition space and study how that prior affects the selection, but do not rank partition priors by their evidence. Tancredi et al. (2020) end their paper on a unified de-duplication framework by proposing that Pitman–Yor-type partition priors be investigated; Pitman–Yor is among the eight priors ranked below. Variational lower bounds on the evidence have been used in microclustering entity resolution (Beraha and Favaro, 2025), for inference within one model rather than for comparison across priors. Pair-level model selection in the Fellegi–Sunter tradition (Fellegi and Sunter, 1969; Larsen and Rubin, 2001) also differs from the comparison made here: it chooses between match and nonmatch models for each record pair, whereas we rank partition priors under one joint model.

For eight partition priors (uniform, Dirichlet process, Pitman–Yor, the blink-style uniform-labelings prior, two ESC-NB settings, and two plug-in-μ ESC-D settings) we computed the log evidence, the evidence ranking, the pairwise Bayes factors, and the entity-count posterior $P(K \mid \text{data}, \text{blocking})$ on RLdata500, RLdata10000, and SHIW0810, with the hit-and-miss likelihood of Section S8. The block sums are computed in double precision by the enumeration method of that section; on RLdata500 the sums under plug-in distortion were also computed in exact integer arithmetic, and the two agree to better than 10^{-11} in $\log Z$. The primary hyperparameters are those of Section S8 (Dirichlet process $\alpha = 1$, Pitman–Yor $\theta = 1$, $\sigma = 1/2$) with the ESC settings listed in Table S7. The sensitivity analysis varies the Dirichlet-process concentration over $\alpha \in \{1/4, 1/2, 2, 4\}$, Pitman–Yor $\theta \in \{1/2, 2\}$, ESC-NB $p \in \{1/5, 1/2, 2/3\}$ at each r, and blink $N_{\max} = 2N$, and, on the RLdata corpora only, Pitman–Yor $\sigma \in \{1/4, 3/4\}$ and ESC-NB $r = 3$. The three corpora hold 450, 9,000, and 28,584 true entities. Even an error-free linkage within blocks gives an entity count above the truth whenever a true pair is split across blocks, since records in different blocks are never linked. On RLdata500 this error-free count, the blocking floor, is 458 against a truth of 450 (8 of 50 true pairs split); on RLdata10000 it is 9,407 against 9,000 (407 of 1,000 split). On SHIW0810 the corpus-level floor is 29,731 against 28,584 (1,147 of 11,159 true links split). The analysis of that corpus is, however, restricted to the 8,609 blocks of at most 14 records, which cover 34,275 of the 39,743 records (86.2%), and the 302 larger blocks are excluded. Within the analyzed blocks the floor is 25,847 against a truth of 24,933.

prior	RLdata500 log Z	rank	RLdata10000 log Z	rank	SHIW0810 log Z	rank
DP	-8713.702	1	-182886.383	1	-207956.408	1
uniform	-8784.545	2	-183070.124	2	-223441.317	2
blink	-9186.090	3	-193468.189	3	-392854.495	4
ESC-D (plug-in μ_{D2})	-9204.512	4	-194929.654	4	-385221.903	3
ESC-NB(r=1, p=$\frac{1}{2}$)	-9278.918	5	-196007.008	5	-397455.106	5
ESC-D (plug-in μ_{D1})	-9278.918	6	-196007.008	6	-399181.828	6
ESC-NB(r=2, p=$\frac{1}{3}$)	-9283.285	7	-196073.352	7	-400688.491	8
PY	-9316.659	8	-196538.344	8	-400208.864	7

Table S7: Log evidence (natural log, after the global-K convolution) and rank of eight partition priors on the three corpora, under the primary Beta-integrated likelihood and the primary hyperparameters. The ESC-D priors are plug-in-μ ESC priors truncated at cluster size 6; the truncation has no effect on RLdata10000, leaves the printed decimals unchanged on RLdata500, and has a large effect on SHIW0810 (see text).

Under the primary Beta-integrated likelihood the Dirichlet process has the highest evidence on all three corpora, and the priors built for microclustering rank below it and below the uniform prior (Table S7). The largest pairwise $\log_{10}$ Bayes factors are 261.86 (Dirichlet process against Pitman–Yor on RLdata500), 5,928.97 (Dirichlet process against Pitman–Yor on RLdata10000), and 83,702.48 (Dirichlet process against ESC-NB(r=2, p=$\frac{1}{3}$) on SHIW0810). The ESC-D priors in the table are plug-in-μ ESC with support $s \leq 6$, so that cluster sizes of 7 and above have prior mass zero. The truncation has no effect on RLdata10000, where no block exceeds six records, and leaves the printed decimals unchanged on RLdata500. On SHIW0810, whose analyzed blocks reach 14 records, ESC-D(μ_{D1}) lies about 750 $\log_{10}$ units below its untruncated counterpart ESC-NB(r=1, p=$\frac{1}{2}$); the Dirichlet-marginalized ESC-D was not computed. Where the ESC priors rank relative to the Dirichlet process and the uniform prior depends on the likelihood specification. They have lower evidence than both under the primary Beta-integrated likelihood on all three corpora and under Beta$(1, 9)$ distortion on the two RLdata corpora. However, under plug-in distortion on RLdata500 all four ESC priors have higher evidence than the Dirichlet process ($\log_{10}$ Bayes factors 22.98 to 44.87) and ESC-D(μ_{D2}) has higher evidence than uniform ($\log_{10}$ Bayes factor 2.98). In the sensitivity analysis, lowering the Dirichlet-process concentration α raises the evidence of the Dirichlet process further above that of the uniform prior on every corpus, and the ESC priors stay below both the Dirichlet process and the uniform prior at every value of p. At larger concentrations the uniform prior overtakes the Dirichlet process under the primary likelihood, at $\alpha = 2$ and $\alpha = 4$ on RLdata10000 and at $\alpha = 4$ on RLdata500. On SHIW0810 the Dirichlet process has higher evidence than the uniform prior at every value of α computed.

On the smallest corpus, RLdata500, the prior with the highest evidence changes under both alternative distortion models. At the primary hyperparameters it is the blink-style prior under plug-in distortion and the uniform prior under Beta$(1, 9)$ distortion, with $\log_{10}$ Bayes factors of 53.61 and 11.35 against the Dirichlet process. In the sensitivity analysis DP(α=4) outranks the blink-style prior under plug-in distortion ($\log_{10}$ Bayes factor 2.84). On RLdata10000 and SHIW0810, at the primary hyperparameters, the Dirichlet process has the highest evidence of the eight priors under every likelihood specification computed. The Pitman–Yor σ variants, ESC-NB r=3, and the Beta$(1, 9)$ distortion were computed on the RLdata corpora only, so the SHIW0810 results have not been checked against those variations.

On the two RLdata corpora, under the primary likelihood and hyperparameters, the prior whose entity-count posterior $P(K \mid \text{data}, \text{blocking})$ lies nearest the blocking floor is Pitman–Yor, with posterior mean 442.2 on RLdata500 against a floor of 458 and 9,210.6 on RLdata10000 against 9,407. Pitman–Yor also has the lowest evidence of the eight priors on both corpora (Table S7). None of the eight posteriors reaches the floor on either corpus: each puts posterior probability at least 0.99999 on entity counts below the floor, so that every prior merges more records than an error-free within-block linkage would, Pitman–Yor by the smallest margin. On SHIW0810 the posteriors fall further below the floor of 25,847, with entity counts within the analyzed blocks ranging from 9,023 (the Dirichlet process, the prior with the highest evidence) to 18,226 (the blink-style prior). Figure S2 shows the posteriors with truth and floor overlaid. The evidence of a prior measures how well the resulting joint model predicts the observed fields under the assumed likelihood. When that likelihood is misspecified, the prior with the highest evidence can therefore be one whose entity count is far from the truth, and computing the evidence without Monte Carlo error does not change this. The log evidences in Table S7 are nevertheless reference values, on these corpora under these blockings and this likelihood, for approximate evidence computations such as variational bounds (Beraha and Favaro, 2025).

S10 Mutational-signature evidence at gene-panel scale

Clinical gene panels report a handful of somatic mutations per tumor, each sorted into one of 96 trinucleotide channels, and mutational signatures are inferred from those few counts. Signature SBS3 marks homologous-recombination deficiency, which predicts response to PARP-inhibitor therapy, and the caller SigMA detects it in targeted-panel data (Gulhan et al., 2019). SigMA's clinical validation study (Batalini et al., 2022) admits samples with as few as five single-nucleotide variants, at a classifier threshold tuned for a 10% false-positive rate and 74% sensitivity. In simulations of how well targeted panels reflect whole-exome signatures, Kodama et al. (2025) found that cancers with few mutations need large panels. Panel-scale presence calling, in SigMA and in refitting tools alike, is threshold-based, with cutpoints tuned on classifier or refitting scores (Lawrence et al., 2021). Such a cutpoint is a proxy for the Bayes factor for signature presence, the ratio of the marginal likelihoods of the panel counts under the models with and without the signature. We compute that Bayes factor directly in this section, for panel-sized readouts, where a readout is one vector of mutation counts over the 96 channels. In the notation of Table 6 of the main paper this application has $g = 10$ against 9 fixed components, $k = 96$ channels, and only $N = 5$ to 30 mutations per readout.

We fix the component profiles at the published COSMIC v3.3.1 signature profiles (GRCh37) over the $k = 96$ trinucleotide mutation channels and integrate out only the mixture weights. For a panel count vector u with $N = \sum_v u_v$ mutations, signature profiles $p_{a,v}$, and weights $w \sim \mathrm{Dir}(\beta)$, the evidence is

$$Z(u; p, \beta) \;=\; \mathbb{E}_{w\sim\mathrm{Dir}(\beta)}\Big[\prod_{v=1}^{96}\Big(\sum_a w_a\, p_{a,v}\Big)^{u_v}\Big], \tag{S10.1}$$

which is the marginal probability of the ordered sequence of mutations; the multinomial coefficient that converts it to the marginal probability of the count vector u cancels in every ratio

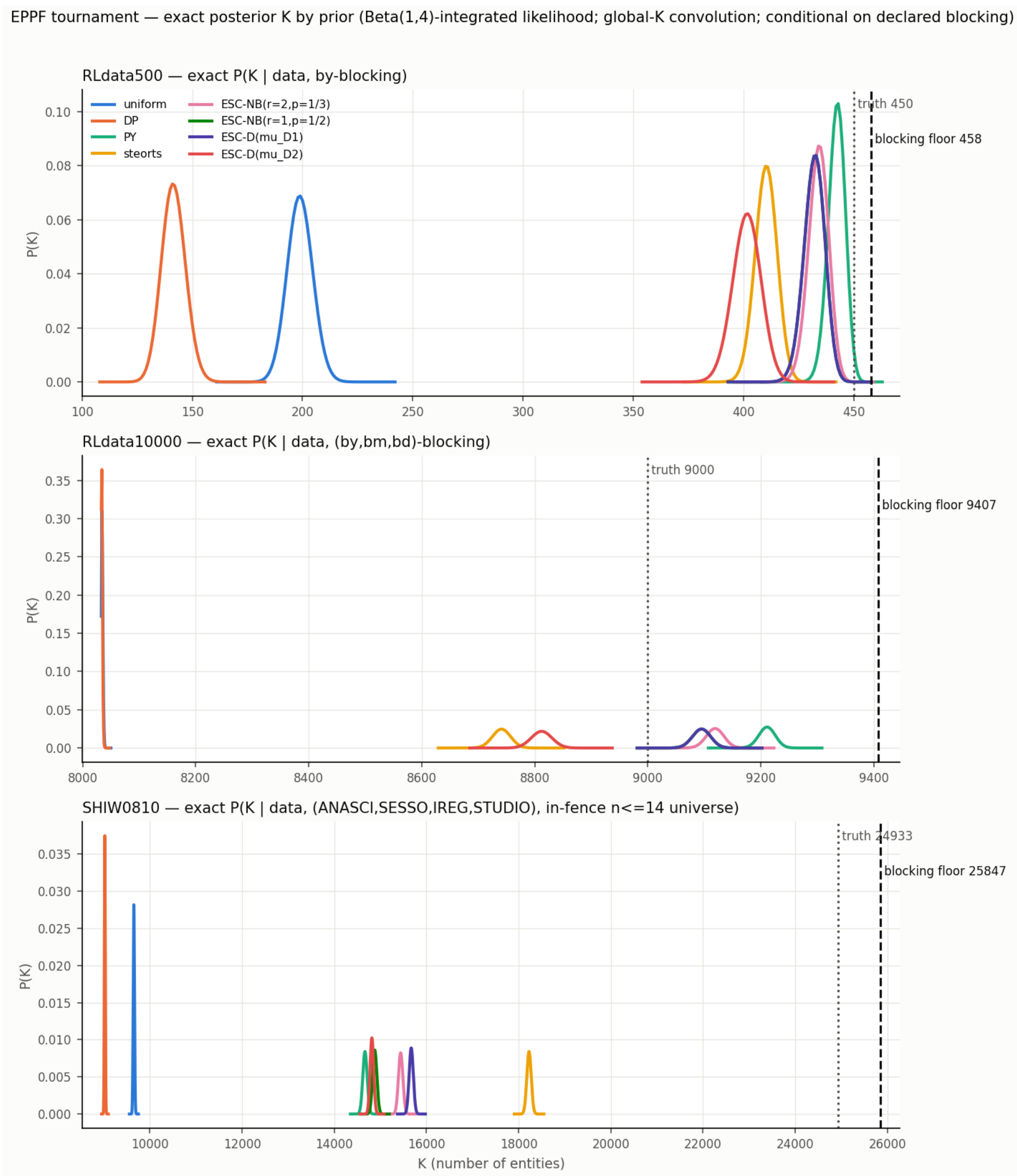

Figure S2: Entity-count posteriors $P(K \mid \text{data}, \text{blocking})$ for the eight partition priors on the three corpora, with the truth and the blocking floor marked, under the primary likelihood and hyperparameters. On every corpus the priors disagree by far more than their posterior widths, and every posterior lies below the floor.

used below. The presence Bayes factor for SBS3 is the ratio of two such evidences, $\mathrm{BF}(u) = Z(u; p_S, \beta) \,/\, Z(u; p_{S\setminus \mathrm{SBS3}}, \beta)$, with S the ten breast-cancer signatures documented in the SigMA paper's open supplement (SBS1, 2, 3, 5, 8, 13, 17a, 17b, 18, 30) and $S \setminus \mathrm{SBS3}$ the other nine. The prior is symmetric on each model's own weight simplex: Dir(1) for all results below, with Dir(1/2) and Dir$(1/g)$, g being the model's own component count, used to test sensitivity to the prior. Marginalization over Dirichlet weights with fixed components is Maennel's integral (Maennel, 2020). The Bayes factors are computed in floating point and agree with an exact-rational implementation to within 4×10^{-15} in $\log_{10}$BF on test cases; for 704 readouts at $N \leq 22$ the Bayes factors recomputed in exact rational arithmetic match the floating-point values to ten decimals (no readout at $N = 30$ was recomputed exactly). On the simulated panel readouts described in the next paragraph, replacing the Dir(1) prior by Dir(1/2) shifts $\log_{10}$BF by a median of 0.012 at $N = 5$, rising to 0.068 at $N = 30$, and by 0.384 on the single most affected readout. Under Dir($1/g$) the corresponding medians are 0.035 to 0.175 over the same range of N and the largest change on a single readout is 1.542. The COSMIC v3.4 revision of the profiles changes $\log_{10}$BF by less than 10^{-6} on the $N = 10$ readouts. Existing Bayesian signature tools such as sigfit (Gori and Baez-Ortega, 2020) sample exposures by MCMC and call presence from posterior intervals, and to our knowledge no published work computes signature-presence evidence, even by bridge sampling. The only formal significance test for signature presence that we are aware of is the likelihood-ratio SignaturePresenceTest of mSigAct (Ng et al., 2017; Jiang et al., 2025).

The test bed is 698 breast-cancer whole genomes from open-access tiers, 484 from the 560-genome cohort of Nik-Zainal et al. (2016) and 214 from the pan-cancer compendium (Alexandrov et al., 2020). Each has a whole-genome reference label: the SigProfiler SBS3 attribution for all 698 and, for 439 of the 484 cohort samples, the CHORD homologous-recombination (HR) call (Nguyen et al., 2020) (the pan-cancer samples cannot be joined to CHORD from open data). We call a sample truth-negative if it is HR-proficient with whole-genome SBS3 exposure zero (195 samples) and truth-positive if it is HR-deficient with SBS3 present (66 samples). From each genome we drew panel-sized mutation subsets at $N = 5, 10, 15, 20, 30$ by multinomial sampling from the genome's 96-channel frequencies, with five independent draws under each of two weighting schemes, or 6,980 readouts per panel size. The first scheme uses the channel frequencies as they stand; the other reweights each channel by the trinucleotide abundance of the panel footprint relative to that of the genome (factors 0.95 to 1.23). A second set of readouts uses real gene footprints: each cohort tumor's own mutations are intersected with the full gene spans of the 468-gene MSK-IMPACT panel (46.6 Mb). These spans are a superset of the regions the panel actually captures, so the footprint counts, with a median of 51 mutations per tumor, are higher than a real panel would report. The Bayes factors were evaluated for footprint readouts of up to $N = 40$ mutations, so the 141 real-footprint readouts in the comparison are the tumors with 1 to 31 footprint mutations (none has 32 to 40, and 290 of the 484 exceed 40). We call a Bayes factor of at least 3 support for the signature and a Bayes factor of at least 20 strong support, the conventional positive and strong thresholds (Kass and Raftery, 1995).

N	max $\log_{10}$BF observed	pure-SBS3 draw max	BF ≥ 20 reached
5	0.900	0.495	no
10	1.311	0.516	yes
15	1.853	1.066	yes
20	1.993	1.775	yes
30	2.701	2.287	yes

Table S8: Largest $\log_{10}$ Bayes factor for SBS3 presence among the 6,980 panel-sized readouts drawn from real tumors at each panel size N, the largest among three multinomial draws from the pure SBS3 profile, and whether any readout reached BF = 20 ($\log_{10} 20 = 1.30$). These are maxima over the simulated readouts; the maximum over all possible five-mutation readouts is BF = 50.27.

At five mutations, the minimum the clinical caller SigMA accepts, the largest Bayes factor among the 6,980 panel-sized readouts drawn from real tumors was 8, well short of 20, and draws from the pure SBS3 profile itself reached only about 3. Over all $\binom{100}{5} = 75{,}287{,}520$ five-mutation count vectors, enumerated in full, the maximum Bayes factor under the same model and prior is 50.27 ($\log_{10} \mathrm{BF} = 1.70135$). It is attained at a unique readout, all five mutations in the channel G[C>G]A, where the SBS3 profile has probability 0.0127, against 0.0047 for SBS5, the largest value among the other nine signatures. Strong support from a five-mutation readout is therefore possible in principle, although none of the readouts drawn from real tumors at that size came close to BF $=$ 20. Even at twenty mutations, only one of the three pure-signature draws gave a Bayes factor above 20 (Table S8). On the real-footprint readouts the maximum is $\log_{10}\mathrm{BF} = 1.392$ at $N = 27$. These maxima apply to inference from the 96-channel count vector alone, under the fixed signature set and the Dirichlet weight prior; classifiers that use features beyond the count vector, as SigMA does (Gulhan et al., 2019), are not bounded by them.

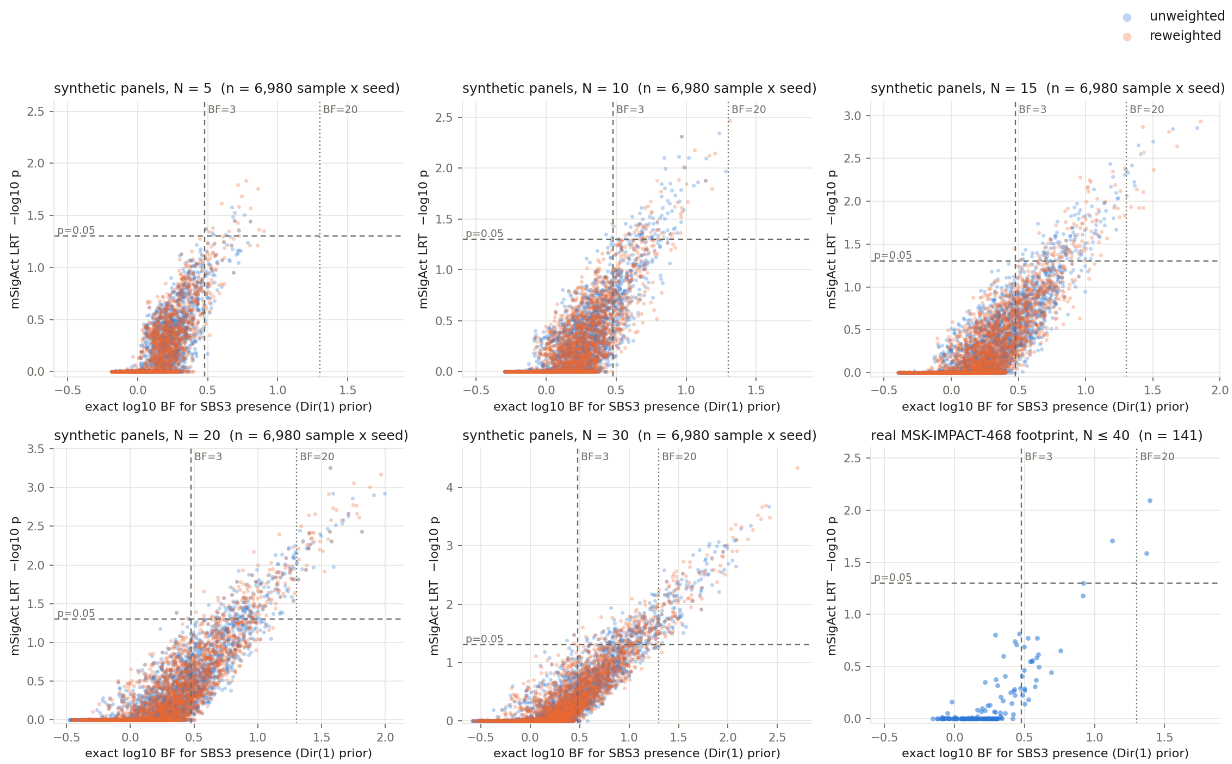

Figure S3: $\log_{10}$ Bayes factor for SBS3 presence (horizontal) against $-\log_{10} p$ of the likelihood-ratio test (vertical), one panel per simulated panel size (blue unweighted, orange reweighted draws) and one for the 141 real-footprint readouts, with lines at BF $=$ 3, BF $=$ 20 and $p = 0.05$. The two methods rank readouts alike, but the test calls only the far tail of the region with BF $\geq$ 3.

When the Bayes factor and the mSigAct likelihood-ratio test are compared readout by readout, the test at $p < 0.05$ calls far fewer readouts than reach BF $\geq$ 3. At thirty mutations, 522 readouts have $p < 0.05$, about a quarter of the 2,047 that have BF $\geq$ 3. On the 330 truth-positive readouts at $N = 30$, BF $\geq$ 3 is reached on 67.6% and 71.5% under the two weighting schemes and the test calls 25.2% and 25.5%. On the 975 truth-negative readouts the corresponding rates are 13.5% and 14.1% against 2.8% and 2.7%. At five mutations, BF ≥ 3 is reached on 5.2% and 6.7% of truth-positive readouts and on 0.9% and 0.6% of truth-negative ones. The two methods agree in direction: among all simulated readouts only 12 of 1,263 test calls have BF below 3, and the rank correlation between $\log_{10}$BF and $-\log_{10} p$ runs from 0.546 to 0.801 across the simulated strata (0.644 on the real-footprint readouts). They appear to differ mainly in sensitivity, by a factor near three at $N = 30$ (Figure S3). As implemented, the test refers twice the log likelihood ratio to a chi-square with one degree of freedom, although the null hypothesis of zero SBS3 exposure lies on the boundary of the alternative's parameter space and no boundary correction is applied; we ran the code as published. In the rates above, the two methods are compared only at their conventional operating points, BF ≥ 3 and $p < 0.05$. We have not compared the discrimination of the two methods, and the reference labels are whole-genome rather than panel-derived. The computation uses only open materials: the published COSMIC signature profiles, the open-access genomes and papers, and the GPL-3 mSigAct test; SigMA enters only through the published operating point of its clinical validation (Batalini et al., 2022).

References

L. B. Alexandrov et al., "The repertoire of mutational signatures in human cancer," Nature **578** (2020) 94–101.

M. Aoyagi, "A Bayesian learning coefficient of generalization error and Vandermonde matrix-type singularities," Comm. Statist. Theory Methods **39** (2010) 2667–2687.

R. Arnold, Y. Hayakawa and P. Yip, "Capture–recapture estimation using finite mixtures of arbitrary dimension," Biometrics **66** (2010) 644–655.

S. Baillargeon and L.-P. Rivest, "Rcapture: Loglinear models for capture–recapture in R," J. Statistical Software **19**(5) (2007) 1–31.

K. Barger and J. Bunge, "Objective Bayesian estimation for the number of species," Bayesian Analysis **5** (2010) 765–785.

F. Batalini, D. C. Gulhan, V. Mao, A. Tran, M. Polak, N. Xiong, et al., "Mutational signature 3 detected from clinical panel sequencing is associated with responses to olaparib in breast and ovarian cancers," Clin. Cancer Res. **28** (2022) 4714–4723.

M. Beraha and S. Favaro, "Large-scale entity resolution via microclustering Ewens–Pitman random partitions," arXiv:2507.18101 (2025).

B. Betancourt, G. Zanella and R. C. Steorts, "Random partition models for microclustering tasks," J. Amer. Statist. Assoc. **117** (2022) 1215–1227, arXiv:2004.02008.

O. Binette and R. C. Steorts, "(Almost) all of entity resolution," Science Advances **8** (2022) eabi8021.

A. Bostan, P. Lairez and B. Salvy, "Multiple binomial sums," J. Symbolic Comput. **80** (2017) 351–386.

A. Bouchard-Côté, A. Doucet and A. Roth, "Particle Gibbs split-merge sampling for Bayesian inference in mixture models," J. Mach. Learn. Res. **18**(28) (2017) 1–39, arXiv:1508.02663.

T. V. Buttrey, "The denarii of P. Crepusius and Roman Republican mint organization," American Numismatic Society Museum Notes **21** (1976) 67–108.

T. V. Buttrey, "Calculating ancient coin production: I. Facts and fantasies," Numismatic Chronicle **153** (1993) 335–351.

L. Carbone and L. Yarrow, "The aftermath of the First Mithridatic War and Sulla's dictatorship: Some preliminary historical analyses using the 'Roman Republican Die Project'," Revue belge de Numismatique **166** (2020) 285–333.

G. Casella, E. Moreno and F. J. Girón, "Cluster analysis, model selection, and prior distributions on models," Bayesian Analysis **9** (2014) 613–658.

B. J. Castledine, "A Bayesian analysis of multiple-recapture sampling for a closed population," Biometrika **68** (1981) 197–210.

J. B. Copas and F. J. Hilton, "Record linkage: statistical models for matching computer records," J. Roy. Statist. Soc. A **153** (1990) 287–320.

P. De Blasi, S. Favaro, A. Lijoi, R. H. Mena, I. Prünster and M. Ruggiero, "Are Gibbs-type priors the most natural generalization of the Dirichlet process?," IEEE Trans. Pattern Anal. Mach. Intell. **37** (2015) 212–229.

F. de Callataÿ, "Calculating ancient coin production: Seeking a balance," Numismatic Chronicle **155** (1995) 289–311.

G. Di Benedetto, F. Caron and Y. W. Teh, "Nonexchangeable random partition models for microclustering," Ann. Statist. **49** (2021) 1931–1957, arXiv:1711.07287.

R. M. Dorazio and R. A. Erickson, "eDNAoccupancy: An R package for multiscale occupancy modelling of environmental DNA data," Mol. Ecol. Resour. **18** (2018) 368–380.

M. Drton and M. Plummer, "A Bayesian information criterion for singular models," J. R. Stat. Soc. Ser. B **79** (2017) 323–380 (with discussion).

W. R. Edwards and L. Eberhardt, "Estimating cottontail abundance from livetrapping data," J. Wildlife Management **31** (1967) 87–96.

W. W. Esty, "How to estimate the original number of dies and the coverage of a sample," Numismatic Chronicle **166** (2006) 359–364.

W. W. Esty, "The geometric model for estimating the number of dies," in F. de Callataÿ (ed.), *Quantifying Monetary Supplies in Greco-Roman Times*, Edipuglia, Bari (2011) 43–58.

I. P. Fellegi and A. B. Sunter, "A theory for record linkage," J. Amer. Statist. Assoc. **64** (1969) 1183–1210.

T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," Ann. Statist. **1** (1973) 209–230.

S. E. Fienberg, M. S. Johnson and B. W. Junker, “Classical multilevel and Bayesian approaches to population size estimation using multiple lists,” J. R. Stat. Soc. Ser. A **162** (1999) 383–405.

M. Fortini, B. Liseo, A. Nuccitelli and M. Scanu, “On Bayesian record linkage,” Research in Official Statistics **4** (2001) 185–198.

K. Gori and A. Baez-Ortega, “sigfit: flexible Bayesian inference of mutational signatures,” bioRxiv 372896 (2020), doi:10.1101/372896.

D. C. Gulhan, J. J.-K. Lee, G. E. M. Melloni, I. Cortés-Ciriano and P. J. Park, “Detecting the mutational signature of homologous recombination deficiency in clinical samples,” Nat. Genet. **51** (2019) 912–919.

N. Jiang, Y. Wu and S. G. Rozen, “Benchmarking 13 tools for mutational signature attribution, including a new and improved algorithm,” Brief. Bioinform. **26** (2025) bbaf042.

J. E. Johndrow, K. Lum and D. B. Dunson, “Theoretical limits of microclustering for record linkage,” Biometrika **105** (2018) 431–446, arXiv:1703.04955.

M. Jones, E. Matechou, D. Cole, A. Diana, J. Griffin, S. Peixoto, L. Lawson Handley and A. Buxton, “More than presence-absence; modelling (e)DNA concentration across time and space from qPCR survey data,” J. Stat. Theory Pract. **19** (2025) 68.

R. E. Kass and A. E. Raftery, “Bayes factors,” J. Amer. Statist. Assoc. **90** (1995) 773–795.

J. Kim, “Marginal likelihoods for finite-support Huber contamination,” arXiv:2605.26723 (2026).

R. King and S. P. Brooks, “On the Bayesian analysis of population size,” Biometrika **88** (2001) 317–336.

K. Kodama, Y. Ling, H. Ichikawa, T. Wakai and S. Okuda, “Evaluation of potential of targeted sequencing through mutational signature simulation,” PLoS One **20** (2025) e0326071.

J. J. Lahoz-Monfort, G. Guillera-Arroita and R. Tingley, “Statistical approaches to account for false-positive errors in environmental DNA samples,” Mol. Ecol. Resour. **16** (2016) 673–685.

M. D. Larsen and D. B. Rubin, “Iterative automated record linkage using mixture models,” J. Amer. Statist. Assoc. **96** (2001) 32–41.

L. Lawrence, C. A. Kunder, E. Fung, H. Stehr and J. Zehnder, “Performance characteristics of mutational signature analysis in targeted panel sequencing,” Arch. Pathol. Lab. Med. **145** (2021) 1424–1431.

S. Lin, “Ideal-theoretic strategies for asymptotic approximation of marginal likelihood integrals,” arXiv:1003.5338 (2010); J. Algebraic Statistics **8** (2017).

S. Lin, B. Sturmfels and Z. Xu, “Marginal likelihood integrals for mixtures of independence models,” J. Mach. Learn. Res. **10** (2009) 1611–1631, [arXiv:0805.3602].

W. A. Link, “Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities,” Biometrics **59** (2003) 1123–1130.

L. Lipshitz, “The diagonal of a D-finite power series is D-finite,” J. Algebra **113** (1988) 373–378.

D. Madigan and J. C. York, “Bayesian methods for estimation of the size of a closed population,” Biometrika **84** (1997) 19–31.

H. Maennel, “Exact marginal inference in Latent Dirichlet Allocation,” arXiv:2004.00115 (2020).

S. Mano, “Partition structure and the A-hypergeometric distribution associated with the rational normal curve,” Electron. J. Statist. **11** (2017) 4452–4487, [arXiv:1607.03569].

D. Manrique-Vallier, “Bayesian population size estimation using Dirichlet process mixtures,” Biometrics **72** (2016) 1246–1254.

N. G. Marchant, A. Kaplan, D. N. Elazar, B. I. P. Rubinstein and R. C. Steorts, “d-blink: Distributed end-to-end Bayesian entity resolution,” J. Comput. Graph. Statist. **30** (2021) 406–421, arXiv:1909.06039.

N. G. Marchant, B. I. P. Rubinstein and R. C. Steorts, “Bayesian graphical entity resolution using exchangeable random partition priors,” J. Surv. Stat. Methodol. **11** (2023) 569–596, arXiv:2301.02962.

C. M. Martel, M. Sutter, R. M. Dorazio and A. P. Kinziger, “Using environmental DNA and occupancy modelling to estimate rangewide metapopulation dynamics,” Mol. Ecol. **30** (2021) 3340–3354.

E. Matechou, G. K. Nicholls, B. J. T. Morgan, J. A. Collazo and J. E. Lyons, “Bayesian analysis of Jolly–Seber type models: incorporating heterogeneity in arrival and departure,” Environ. Ecol. Stat. **23** (2016) 531–547 [arXiv:1512.05170].

B. S. McVeigh, B. T. Spahn and J. S. Murray, “Scaling Bayesian probabilistic record linkage with post-hoc blocking: an application to the California Great Registers,” arXiv:1905.05337 (2019).

G. Meshcheryakov, “Exact computation of posterior distribution of mixture weights in hierarchical Bayesian models,” arXiv:2607.05692 (2026).

J. W. Miller, B. Betancourt, A. Zaidi, H. Wallach and R. C. Steorts, “Microclustering: When the cluster sizes grow sublinearly with the size of the data set,” arXiv:1512.00792 (2015).

D. Mirzaee, S. A. Malekpour and A. Kalirad, “The exact hypergeometric posterior method for accurate inference of population size from mark-recapture data,” Bull. Math. Biol. **88** (2026) 128.

A. Mozeika and A. C. C. Coolen, “Replica analysis of Bayesian data clustering,” J. Phys. A **53** (2020) 034001, [arXiv:1810.02627].

A. W. T. Ng et al., “Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia,” Sci. Transl. Med. **9** (2017) eaan6446.

L. Nguyen, J. W. M. Martens, A. Van Hoeck and E. Cuppen, “Pan-cancer landscape of homologous recombination deficiency,” Nat. Commun. **11** (2020) 5584.

S. Nik-Zainal et al., "Landscape of somatic mutations in 560 breast cancer whole-genome sequences," Nature **534** (2016) 47–54.

A. Orlitsky, N. P. Santhanam, and J. Zhang, "Universal compression of memoryless sources over unknown alphabets," IEEE Trans. Inf. Theory **50** (2004) 1469–1481.

J. Pitman and M. Yor, "The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator," Ann. Probab. **25** (1997) 855–900.

S. Pledger, "Unified maximum likelihood estimates for closed capture–recapture models using mixtures," Biometrics **56** (2000) 434–442.

J. A. Royle, "Site occupancy models with heterogeneous detection probabilities," Biometrics **62** (2006) 97–102.

J. A. Royle and W. A. Link, "Generalized site occupancy models allowing for false positive and false negative errors," Ecology **87** (2006) 835–841.

D. Rusakov and D. Geiger, "Asymptotic model selection for naive Bayesian networks," J. Mach. Learn. Res. **6** (2005) 1–35.

M. Sadinle, "Detecting duplicates in a homicide registry using a Bayesian partitioning approach," Ann. Appl. Statist. **8** (2014) 2404–2434.

M. C. Schmelzle and A. P. Kinziger, "Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species," Mol. Ecol. Resour. **16** (2016) 895–908.

H. S. Seung, H. Sompolinsky, and N. Tishby, "Statistical mechanics of learning from examples," Phys. Rev. A **45** (1992) 6056–6091.

P. J. Smith, "Bayesian analyses for a multiple capture–recapture model," Biometrika **78** (1991) 399–407.

R. Sollmann, M. J. Eaton, W. A. Link, P. Mulondo, S. Ayebare, S. Prinsloo, A. J. Plumptre and D. S. Johnson, "A Bayesian Dirichlet process community occupancy model to estimate community structure and species similarity," Ecol. Appl. **31** (2021) e02249.

A. J. Stam, "Statistical problems in ancient numismatics," Statistica Neerlandica **41** (1987) 151–174.

R. C. Steorts, "Entity resolution with empirically motivated priors," Bayesian Analysis **10** (2015) 849–875, arXiv:1409.0643.

W. Szpankowski and M. J. Weinberger, "Minimax pointwise redundancy for memoryless models over large alphabets," IEEE Trans. Inf. Theory **58** (2012) 4094–4104.

Y. Tachibana, Y. Goto, T. Koyama and N. Takayama, "Holonomic gradient method for two-way contingency tables," Algebraic Statistics **11** (2020) 125–153, [arXiv:1803.04170].

A. Tancredi and B. Liseo, "A hierarchical Bayesian approach to record linkage and population size problems," Ann. Appl. Statist. **5** (2011) 1553–1585, arXiv:1011.2649.

A. Tancredi, R. Steorts and B. Liseo, "A unified framework for de-duplication and population size estimation (with discussion)," Bayesian Analysis **15** (2020) 633–682.

D. Turek, C. Wehrhahn and O. Gimenez, "Bayesian non-parametric detection heterogeneity in ecological models," Environ. Ecol. Stat. **28** (2021) 355–381, arXiv:2007.10163.

T. Watanabe and S. Watanabe, "Asymptotic behavior of Bayesian generalization error in multinomial mixtures," arXiv:2203.06884 (2022).

G. Zanella, B. Betancourt, H. Wallach, J. Miller, A. Zaidi and R. C. Steorts, "Flexible models for microclustering with application to entity resolution," Advances in Neural Information Processing Systems **29** (2016), arXiv:1610.09780.