

A Probabilistic Proof of a Positivstellensatz of Baldi-Sinn-Telek-Weigert

17 Oct 2026

Abstract

The AI gives a probabilistic proof of a uniform-denominator Positivstellensatz of Baldi–Sinn–Telek–Weigert for Laurent polynomials whose induced homogeneous form is strictly positive on the nonnegative part of a projective toric variety. Only this abstract was human-written. Three classical tools from mathematical statistics and probability theory were used by the AI: the Barndorff-Nielsen completion of discrete exponential families, sufficiency-based invariance under exponential tilting following Halmos–Savage and Van Campenhout–Cover, and the conditional limit theorem (equivalence of ensembles) of Diaconis–Freedman.

1 Exponential Families on Lattice Configurations and Toric Closures

Let $\mathcal{A} = \{a_1, \dots, a_m\} \subset \mathbb{Z}^n$ be a non-empty finite set of lattice points, and let $P := \text{conv}(\mathcal{A}) \subset \mathbb{R}^n$ denote its convex support. For any non-empty face $F \preceq P$, we write $\mathcal{A}_F := \mathcal{A} \cap F$, so that $F = \text{conv}(\mathcal{A}_F)$.

The general theory of natural exponential families generated by a carrier measure on $\mathbb{R}^n$ was systematically developed by Barndorff-Nielsen [2, Chapter 8] and Brown [5, Chapter 1], and is surveyed from the perspective

of graphical models and variational inference by Wainwright–Jordan [24, Section 3]. We recall the standard definition following Brown [5, Chapter 1].

Definition 1.1 (Brown [5, Chapter 1]). Let ν be the counting measure on the finite set $\mathcal{A} \subset \mathbb{Z}^n$. The *natural exponential family* $\mathcal{E}_{\mathcal{A}} = \{P_{\theta} : \theta \in \mathbb{R}^n\}$ generated by ν with natural parameter $\theta \in \mathbb{R}^n$ is the family of probability distributions on $\mathcal{A}$ with probability mass function

$$P_{\theta}(X = a) := \exp(\langle \theta, a \rangle - \Lambda_{\mathcal{A}}(\theta)), \quad a \in \mathcal{A}, \quad (1)$$

where $\Lambda_{\mathcal{A}}(\theta) := \log \sum_{a' \in \mathcal{A}} e^{\langle \theta, a' \rangle}$ is the *cumulant generating function* (or *log-partition function*). Under the multiplicative reparameterization $t = e^{\theta} := (e^{\theta_1}, \dots, e^{\theta_n}) \in \mathbb{R}_{>0}^n$, we have $P_{\theta}(X = a) = t^a / \sum_{a' \in \mathcal{A}} t^{a'}$.

Because $\mathcal{A}$ is finite, the natural parameter space is $\Theta = \mathbb{R}^n$, so $\mathcal{E}_{\mathcal{A}}$ is a full, regular exponential family in the sense of Brown [5, Chapter 1]. By Brown [5, Theorem 2.2], $\Lambda_{\mathcal{A}}$ is real-analytic and convex on $\mathbb{R}^n$, and its gradient and Hessian are the mean vector and covariance matrix of $X \sim P_{\theta}$:

$$\begin{aligned} \mu_{\mathcal{A}}(\theta) &:= \nabla \Lambda_{\mathcal{A}}(\theta) = \mathbb{E}_{P_{\theta}}[X] = \sum_{a \in \mathcal{A}} a P_{\theta}(X = a), \\ \Sigma_{\mathcal{A}}(\theta) &:= \nabla^2 \Lambda_{\mathcal{A}}(\theta) = \text{Cov}_{P_{\theta}}(X). \end{aligned} \quad (2)$$

For each non-empty face $F \preceq P$, let $V_F := \text{span}_{\mathbb{R}}(\mathcal{A}_F - a_0)$ (for any fixed $a_0 \in \mathcal{A}_F$) be the linear subspace parallel to the affine hull of F , of dimension $d_F := \dim(F)$. By Brown [5, Proposition 1.4], the covariance operator $\Sigma_{\mathcal{A}_F}(\theta)$ of the facial exponential family $\mathcal{E}_{\mathcal{A}_F}$ supported on $\mathcal{A}_F$ restricts to a strictly positive definite quadratic form on V_F for every $\theta \in \mathbb{R}^n$.

A foundational result in the theory of exponential families with finite support describes the topological closure $\bar{\mathcal{E}}_{\mathcal{A}}$ of $\mathcal{E}_{\mathcal{A}}$ inside the probability simplex $\Delta_{m-1} = \{p \in \mathbb{R}_{\geq 0}^{\mathcal{A}} : \sum_{a \in \mathcal{A}} p_a = 1\}$ together with the behavior of the mean-value map $\mu : \mathcal{Q} \mapsto \mathbb{E}_{\mathcal{Q}}[X]$. The homeomorphism between the parameter space of a log-linear model and the interior of its marginal polytope originates in the classical theorem of Birch [4] and was established for general regular exponential families by Barndorff-Nielsen [2, Theorem 9.1]

and Brown [5, Theorem 3.6]. The facial stratification of the boundary completion $\overline{\mathcal{E}}_{\mathcal{A}}$ was proved for finite support by Barndorff-Nielsen [2, Theorem 9.2] (where $\overline{\mathcal{E}}_{\mathcal{A}}$ is termed the *aggregate exponential family*; see also Brown [5, Chapter 6], Csiszár–Matúš [8], Geyer [13, Theorem 4], and Rauh–Kahle–Ay [20]). In algebraic statistics, Geiger–Meek–Sturmfels [12, Theorem 3.2] showed that $\overline{\mathcal{E}}_{\mathcal{A}}$ coincides with the nonnegative part $(X_{\mathcal{A}})_{\geq 0} := X_{\mathcal{A}} \cap \Delta_{m-1}$ of the projective toric variety $X_{\mathcal{A}} \subset \mathbb{P}^{m-1}$, under which the mean-value map μ is the algebraic moment map. We state the completion theorem following the formulation of Barndorff-Nielsen [2, Theorem 9.2].

Theorem 1.2 (Barndorff-Nielsen [2, Theorem 9.2]). *Let $\mathcal{E}_{\mathcal{A}}$ be the natural exponential family generated by the counting measure on the finite set $\mathcal{A} \subset \mathbb{Z}^n$ with convex support $P = \text{conv}(\mathcal{A})$, and for each non-empty face $F \preceq P$, let $\mathcal{E}_{\mathcal{A}_F}$ be the natural exponential family generated by the counting measure on $\mathcal{A}_F = \mathcal{A} \cap F$. Define the aggregate exponential family*

$$\overline{\mathcal{E}}_{\mathcal{A}} := \bigsqcup_{F \preceq P} \mathcal{E}_{\mathcal{A}_F} \subset \Delta_{m-1}. \quad (3)$$

Then:

- (i) $\overline{\mathcal{E}}_{\mathcal{A}}$ is the closure of $\mathcal{E}_{\mathcal{A}}$ in the space of probability measures on $\mathcal{A}$ under the weak (equivalently, total variation) topology, and hence is compact.
- (ii) For each non-empty face $F \preceq P$, the mean-value map $\mu_F : \mathcal{E}_{\mathcal{A}_F} \rightarrow \text{relint}(F)$ given by $P_{F,\theta} \mapsto \mathbb{E}_{P_{F,\theta}}[X]$ is a homeomorphism (and a real-analytic diffeomorphism when parameterized by $\theta \in V_F$).
- (iii) The extended mean-value map

$$\mu : \overline{\mathcal{E}}_{\mathcal{A}} \longrightarrow P = \text{conv}(\mathcal{A}), \quad Q \longmapsto \mathbb{E}_Q[X] = \sum_{a \in \mathcal{A}} Q(a) a, \quad (4)$$

is a bijection and a homeomorphism of compact topological spaces.

2 Probabilistic Representation of $\mathcal{A}$ -Copositivity and Toric Multipliers

For any integer $m \geq 1$ and $u \in \mathbb{Z}^n$, we denote by $m \cdot \mathcal{A} := \{a^{(1)} + \dots + a^{(m)} : a^{(1)}, \dots, a^{(m)} \in \mathcal{A}\}$ the m -fold Minkowski sum of $\mathcal{A}$, and by

$$W_m(u) := [t^u] \left(\sum_{a \in \mathcal{A}} t^a \right)^m = \#\{(a^{(1)}, \dots, a^{(m)}) \in \mathcal{A}^m : a^{(1)} + \dots + a^{(m)} = u\} \quad (5)$$

the number of m -step lattice walks with steps in $\mathcal{A}$ ending at u . Clearly $W_m(u) > 0$ if and only if $u \in m \cdot \mathcal{A}$. Similarly, for each non-empty face $F \preceq P$, we write $W_{m,F}(u) := \#\{(a^{(1)}, \dots, a^{(m)}) \in \mathcal{A}_F^m : a^{(1)} + \dots + a^{(m)} = u\}$.

Lemma 2.1 (Facial Walk Restriction). *Let $F \preceq P$ be a non-empty face and let $m \geq 1$. Then $(m \cdot \mathcal{A}) \cap mF = m \cdot \mathcal{A}_F$, and for every $u \in m \cdot \mathcal{A}_F$,*

$$W_m(u) = W_{m,F}(u). \quad (6)$$

Proof. Since F is a face of $P = \text{conv}(\mathcal{A})$, there exist an inward-pointing normal vector $v_F \in \mathbb{R}^n$ and a scalar $\lambda_F \in \mathbb{R}$ such that $\langle v_F, a \rangle = \lambda_F$ for all $a \in \mathcal{A}_F$ and $\langle v_F, a \rangle > \lambda_F$ for all $a \in \mathcal{A} \setminus \mathcal{A}_F$. For any m -tuple $(a^{(1)}, \dots, a^{(m)}) \in \mathcal{A}^m$ with sum $u = \sum_{j=1}^m a^{(j)}$, we have

$$\langle v_F, u \rangle = \sum_{j=1}^m \langle v_F, a^{(j)} \rangle \geq m\lambda_F,$$

with equality (which holds if and only if $u \in mF$) if and only if $\langle v_F, a^{(j)} \rangle = \lambda_F$ for every $j = 1, \dots, m$, i.e., $a^{(j)} \in \mathcal{A}_F$ for all j . Thus every m -step walk in $\mathcal{A}$ summing to $u \in mF$ uses solely steps in $\mathcal{A}_F$. $\square$

Fix $k \in \mathbb{N}$ and let $f(t) = \sum_{b \in k \cdot \mathcal{A}} c_b t^b \in \mathbb{R}[t_1^{\pm 1}, \dots, t_n^{\pm 1}]$ be a Laurent polynomial whose support is contained in $k \cdot \mathcal{A}$. For each face $F \preceq P$, let $f_F(t) := \sum_{b \in k \cdot \mathcal{A}_F} c_b t^b$ denote the facial truncation of f to kF (analogous to the facial sub-polynomials on the simplex studied by Castle-Powers-Reznick [6] and Dickinson-Povh [10]). Following Baldi-Sinn-Telek-Weigert [1, Section 2.2], f is *strictly $\mathcal{A}$ -copositive* if the homogeneous

form induced by f is strictly positive on the nonnegative projective toric variety $(X_{\mathcal{A}})_{\geq 0}$. By the facial characterization in Baldi–Sinn–Telek–Weigert [1, Proposition 2.7], f is strictly $\mathcal{A}$ -cpositive if and only if $f_F(t) > 0$ for every non-empty face $F \preceq P$ and every $t \in \mathbb{R}_{>0}^n$.

Define the normalized coefficient observable $g : k \cdot \mathcal{A} \rightarrow \mathbb{R}$ by

$$g(b) := \frac{c_b}{W_k(b)}, \quad b \in k \cdot \mathcal{A}. \quad (7)$$

Lemma 2.2 (Unconditional Gibbs Expectation and Uniform Positivity). *For any probability measure $Q \in \bar{\mathcal{E}}_{\mathcal{A}}$, let $X_1, \dots, X_k$ be independent and identically distributed (i.i.d.) random vectors with law Q , and set $S_k := X_1 + \dots + X_k \in k \cdot \mathcal{A}$.*

(i) *If $Q = P_{F,\theta} \in \mathcal{E}_{\mathcal{A}_F}$ for a face $F \preceq P$ and $\theta \in \mathbb{R}^n$ (with $t = e^\theta$), then*

$$\mathbb{E}_{P_{F,\theta}}[g(S_k)] = \frac{f_F(t)}{(\sum_{a \in \mathcal{A}_F} t^a)^k}. \quad (8)$$

(ii) *The Laurent polynomial f is strictly $\mathcal{A}$ -cpositive if and only if there exists a constant $\delta > 0$ such that*

$$\inf_{Q \in \bar{\mathcal{E}}_{\mathcal{A}}} \mathbb{E}_Q[g(S_k)] = \delta > 0. \quad (9)$$

Proof. (i) Under $P_{F,\theta}^{\otimes k}$, any k -tuple $(a^{(1)}, \dots, a^{(k)}) \in \mathcal{A}_F^k$ with sum $b \in k \cdot \mathcal{A}_F$ has probability $t^b / (\sum_{a \in \mathcal{A}_F} t^a)^k$. Summing over all $W_{k,F}(b)$ such k -tuples gives $\mathbb{P}_{P_{F,\theta}}(S_k = b) = W_{k,F}(b) t^b / (\sum_{a \in \mathcal{A}_F} t^a)^k$. By Lemma 2.1, $W_{k,F}(b) = W_k(b)$ for all $b \in k \cdot \mathcal{A}_F$, whence

$$\mathbb{E}_{P_{F,\theta}}[g(S_k)] = \sum_{b \in k \cdot \mathcal{A}_F} \frac{c_b}{W_k(b)} \frac{W_{k,F}(b) t^b}{(\sum_{a \in \mathcal{A}_F} t^a)^k} = \frac{\sum_{b \in k \cdot \mathcal{A}_F} c_b t^b}{(\sum_{a \in \mathcal{A}_F} t^a)^k} = \frac{f_F(t)}{(\sum_{a \in \mathcal{A}_F} t^a)^k}.$$

(ii) Since $\mathcal{A}$ is finite, the map $Q \mapsto \mathbb{E}_Q[g(S_k)] = \sum_{(a^{(1)}, \dots, a^{(k)}) \in \mathcal{A}^k} g(a^{(1)} + \dots + a^{(k)}) \prod_{j=1}^k Q(a^{(j)})$ is a polynomial in the probabilities $(Q(a))_{a \in \mathcal{A}}$ and is therefore continuous on $\bar{\mathcal{E}}_{\mathcal{A}}$. By Theorem 1.2(i) and part (i), $f_F(t) > 0$ for all $F \preceq P$ and $t \in \mathbb{R}_{>0}^n$ if and only if $\mathbb{E}_Q[g(S_k)] > 0$ for all $Q \in \bar{\mathcal{E}}_{\mathcal{A}}$, which by the compactness of $\bar{\mathcal{E}}_{\mathcal{A}}$ in Theorem 1.2(i) is equivalent to (9). $\square$

We now connect the coefficients of the toric multiplier $(\sum_{a \in \mathcal{A}} t^a)^N f(t)$ to conditioning on a sufficient statistic. By the classical factorization criterion for sufficient statistics due to Halmos–Savage [14, Theorem 1] (see also Lehmann–Romano [16, Theorem 2.7.1]), the sample sum $S_{N+k} = \sum_{i=1}^{N+k} X_i$ is a minimal sufficient statistic for the natural parameter θ of the exponential family $\mathcal{E}_{\mathcal{A}}$. As exploited in the exponential tilting method of Esscher [11], Van Campenhout–Cover [23, Equation (9)], and Diaconis–Freedman [9, Section 2], conditioning an i.i.d. sample from P_θ on $S_{N+k} = \gamma$ eliminates θ and reduces to the uniform microcanonical distribution over lattice trajectories summing to γ .

Lemma 2.3 (Microcanonical Conditional Representation of Toric Multipliers). *Let $N \in \mathbb{N}$, and let $X_1, \dots, X_{N+k}$ be i.i.d. random vectors taking values in $\mathcal{A}$ with law $P_\theta \in \mathcal{E}_{\mathcal{A}}$. Set $S_m := X_1 + \dots + X_m$ for $1 \leq m \leq N+k$.*

- (i) *For every $\gamma \in (N+k) \cdot \mathcal{A}$, the conditional distribution of $(X_1, \dots, X_{N+k})$ given $S_{N+k} = \gamma$ is independent of $\theta \in \mathbb{R}^n$ and is the uniform discrete distribution on the set of trajectories $\{(x_1, \dots, x_{N+k}) \in \mathcal{A}^{N+k} : \sum_{i=1}^{N+k} x_i = \gamma\}$.*
- (ii) *For every $\gamma \in (N+k) \cdot \mathcal{A}$ and $b \in k \cdot \mathcal{A}$, the conditional distribution of S_k given $S_{N+k} = \gamma$ satisfies*

$$\begin{aligned} \mathbb{P}(S_k = b \mid S_{N+k} = \gamma) &= \frac{W_k(b)W_N(\gamma - b)}{W_{N+k}(\gamma)} \\ &= \mathbb{P}_{P_\theta}(S_k = b) \frac{\mathbb{P}_{P_\theta}(S_{N+k} - S_k = \gamma - b)}{\mathbb{P}_{P_\theta}(S_{N+k} = \gamma)}, \end{aligned} \quad (10)$$

and the coefficient of t^γ in $(\sum_{a \in \mathcal{A}} t^a)^N f(t)$ satisfies the exact identity

$$\frac{[t^\gamma] \left((\sum_{a \in \mathcal{A}} t^a)^N f(t) \right)}{W_{N+k}(\gamma)} = \mathbb{E}[g(S_k) \mid S_{N+k} = \gamma]. \quad (11)$$

Proof. Under $P_\theta^{\otimes(N+k)}$, the joint probability mass function of $(X_1, \dots, X_{N+k})$ at $(x_1, \dots, x_{N+k}) \in \mathcal{A}^{N+k}$ is

$$\mathbb{P}_{P_\theta}(X_1 = x_1, \dots, X_{N+k} = x_{N+k}) = \exp \left(\left\langle \theta, \sum_{i=1}^{N+k} x_i \right\rangle - (N+k) \Lambda_{\mathcal{A}}(\theta) \right),$$

which depends on $(x_1, \dots, x_{N+k})$ only through $\sum_{i=1}^{N+k} x_i$. Thus all $W_{N+k}(\gamma)$ trajectories in $\mathcal{A}^{N+k}$ summing to γ have equal conditional probability $1/W_{N+k}(\gamma)$ regardless of θ , proving (i). Partitioning the $N+k$ steps into the first k steps summing to b and the last N steps summing to $\gamma - b$ gives the first equality in (10), while evaluating the conditional probability under $P_\theta^{\otimes(N+k)}$ gives the second equality. Finally, expanding $(\sum_{a \in \mathcal{A}} t^a)^N f(t) = (\sum_{u \in N \cdot \mathcal{A}} W_N(u) t^u) (\sum_{b \in k \cdot \mathcal{A}} c_b t^b)$ yields

$$\begin{aligned} \frac{[t^\gamma] \left((\sum_{a \in \mathcal{A}} t^a)^N f(t) \right)}{W_{N+k}(\gamma)} &= \sum_{b \in k \cdot \mathcal{A}} \frac{c_b}{W_k(b)} \frac{W_k(b) W_N(\gamma - b)}{W_{N+k}(\gamma)} \\ &= \mathbb{E}[g(S_k) \mid S_{N+k} = \gamma]. \end{aligned} \quad \square$$

3 Equivalence of Ensembles and the Proof of Theorem A

The asymptotic convergence of the conditional distribution of the first k coordinates $(X_1, \dots, X_k)$ given the sufficient statistic $S_M = \gamma'$ to the i.i.d. canonical exponential family distribution $P_{\theta(\gamma'/M)}^{\otimes k}$ is known in statistical mechanics and information theory as the *equivalence of ensembles* or the *Gibbs conditioning principle*. Pointwise versions of this principle were proved by Martin-Löf [17], Van Campenhout–Cover [23, Theorem 1], Csiszár [7], and Zabell [25] by combining exponential tilting with the multivariate local central limit theorem for lattice random vectors established by Rvaceva [21], Petrov [18], Spitzer [22, Proposition P9, p. 75], and Bhattacharya–Ranga Rao [3, Theorem 22.1]. A quantitative version holding uniformly over compact subsets of the interior of the mean space in total variation distance $\|\mu - \nu\|_{\text{TV}} := \sup_E |\mu(E) - \nu(E)|$ was established by Diaconis–Freedman [9, Theorem 1.1 and Proposition 2.6]. We state the theorem of Diaconis–Freedman [9, Theorem 1.1] specialized to the facial exponential family $\mathcal{E}_{\mathcal{A}_F}$.

Theorem 3.1 (Diaconis–Freedman [9, Theorem 1.1]). *Let $F \preceq P$ be a non-empty face, and let $K \subset \text{relint}(F)$ be a compact subset. For integers $1 \leq k \leq M$ and $\gamma' \in M \cdot \mathcal{A}_F$, let $Q_{F, \gamma', k}$ denote the conditional distribution of $(X_1, \dots, X_k)$*

given $S_M = \gamma'$ under the exponential family $\mathcal{E}_{\mathcal{A}_F}$. Then there exists a constant $C_K < \infty$, depending only on $\mathcal{A}_F$ and K , such that for all $M \geq k$ and all $\gamma' \in M \cdot \mathcal{A}_F$ satisfying $\gamma'/M \in K$,

$$\left\| Q_{F, \gamma', k} - P_{F, \theta(\gamma'/M)}^{\otimes k} \right\|_{\text{TV}} \leq C_K \frac{k}{M}, \quad (12)$$

where $\theta(\gamma'/M) := \mu_F^{-1}(\gamma'/M) \in V_F$ is the unique parameter in V_F satisfying $\mathbb{E}_{P_{F, \theta(\gamma'/M)}}[X_1] = \gamma'/M$.

We now prove the toric extension of the theorem of Pólya [19] and the Laurent polyhedral theorem of Handelman [15], stated as Theorem A in the introduction and as Theorem 3.6 in the text of Baldi–Sinn–Telek–Weigert [1].

Theorem 3.2 (Baldi–Sinn–Telek–Weigert [1, Theorem 3.6]). *Let $\mathcal{A} = \{a_1, \dots, a_m\} \subset \mathbb{Z}^n$ be a finite set and $k \in \mathbb{N}$. Let $f = \sum_{b \in k \cdot \mathcal{A}} c_b t^b$ be a Laurent polynomial with support contained in $k \cdot \mathcal{A}$. Then f is strictly $\mathcal{A}$ -copositive if and only if there exists $N \in \mathbb{N}$ such that $(\sum_{a \in \mathcal{A}} t^a)^N f$ has only nonnegative coefficients and $\text{Newt}\left((\sum_{a \in \mathcal{A}} t^a)^N f\right) = \text{conv}((N+k) \cdot \mathcal{A})$. Moreover, if f is strictly $\mathcal{A}$ -copositive, then for $N \in \mathbb{N}$ large enough, every monomial with exponent in $(N+k) \cdot \mathcal{A}$ appears in $(\sum_{a \in \mathcal{A}} t^a)^N f$ with strictly positive coefficient.*

Proof. The “if” direction is immediate: if $h_N(t) := (\sum_{a \in \mathcal{A}} t^a)^N f(t)$ has nonnegative coefficients and $\text{Newt}(h_N) = \text{conv}((N+k) \cdot \mathcal{A}) = (N+k)P$, then for every non-empty face $F \preceq P$, the facial truncation $(h_N)_F(t) = (\sum_{a \in \mathcal{A}_F} t^a)^N f_F(t)$ is a non-zero Laurent polynomial with nonnegative coefficients, so $f_F(t) > 0$ for all $t \in \mathbb{R}_{>0}^n$.

For the “only if” direction and the “moreover” statement, assume that f is strictly $\mathcal{A}$ -copositive. By Lemma 2.2(ii), $\delta := \inf_{Q \in \bar{\mathcal{E}}_{\mathcal{A}}} \mathbb{E}_Q[g(S_k)] > 0$. By Lemma 2.3(ii), to prove that every monomial with exponent $\gamma \in (N+k) \cdot \mathcal{A}$ appears in $(\sum_{a \in \mathcal{A}} t^a)^N f(t)$ with strictly positive coefficient for all sufficiently large N , it suffices to show that there exists $N_0 \in \mathbb{N}$ such that

$$\min_{\gamma \in (N+k) \cdot \mathcal{A}} \mathbb{E}[g(S_k) \mid S_{N+k} = \gamma] \geq \frac{\delta}{2} > 0 \quad \text{for all } N \geq N_0. \quad (13)$$

Suppose, for the sake of contradiction, that (13) fails. Then there exist a strictly increasing sequence of positive integers $N_j \rightarrow \infty$ and a sequence of

lattice points $\gamma_j \in (N_j + k) \cdot \mathcal{A}$ such that

$$\mathbb{E}[g(S_k) \mid S_{N_j+k} = \gamma_j] < \frac{\delta}{2} \quad \text{for all } j \geq 1. \quad (14)$$

Since $\mu_j := \frac{\gamma_j}{N_j+k} \in P = \text{conv}(\mathcal{A})$ and P is compact, after passing to a subsequence we may assume that

$$\lim_{j \rightarrow \infty} \mu_j = \mu_\infty \in P.$$

Since $P = \bigsqcup_{F \preceq P} \text{relint}(F)$, there is a unique non-empty face $F \preceq P$ such that $\mu_\infty \in \text{relint}(F)$.

Choose an inward-pointing normal vector $v_F \in \mathbb{R}^n$ and scalar $\lambda_F \in \mathbb{R}$ for the face F as in Lemma 2.1, and define the nonnegative facial discrepancy $\psi_F(a) := \langle v_F, a \rangle - \lambda_F$ for $a \in \mathcal{A}$. Then $\psi_F(a) = 0$ for all $a \in \mathcal{A}_F$, while if $\mathcal{A} \setminus \mathcal{A}_F \neq \emptyset$, then $\eta_F := \min_{a \in \mathcal{A} \setminus \mathcal{A}_F} \psi_F(a) > 0$. (If $\mathcal{A} \setminus \mathcal{A}_F = \emptyset$, then $F = P$ and $M_j = 0$ below.)

For each $j \geq 1$, consider the random walk $(X_1, \dots, X_{N_j+k}) \in \mathcal{A}^{N_j+k}$ conditioned on $S_{N_j+k} = \gamma_j$. Let $I_{\text{off}} := \{i \in \{1, \dots, N_j+k\} : X_i \notin \mathcal{A}_F\}$ be the random set of indices of off-face steps, let $M_j := |I_{\text{off}}|$ be its cardinality, and let $U_j := \sum_{i \in I_{\text{off}}} X_i$ be the sum of the off-face steps. On the conditioning event $\{S_{N_j+k} = \gamma_j\}$, we have the deterministic inequality

$$\eta_F M_j \leq \sum_{i=1}^{N_j+k} \psi_F(X_i) = \langle v_F, \gamma_j \rangle - (N_j+k)\lambda_F = (N_j+k)(\langle v_F, \mu_j \rangle - \lambda_F).$$

Since $\mu_j \rightarrow \mu_\infty \in F$ and $\langle v_F, \mu_\infty \rangle = \lambda_F$, the deterministic upper bound

$$m_j^{\max} := \frac{N_j+k}{\eta_F} (\langle v_F, \mu_j \rangle - \lambda_F) = o(N_j) \quad (j \rightarrow \infty) \quad (15)$$

satisfies $0 \leq M_j \leq m_j^{\max}$ and $\|U_j\| \leq m_j^{\max} \max_{a \in \mathcal{A}} \|a\| = o(N_j)$ for every trajectory in $\{S_{N_j+k} = \gamma_j\}$.

Let $E_j := \{I_{\text{off}} \cap \{1, \dots, k\} = \emptyset\}$ be the event that the first k steps $X_1, \dots, X_k$ all lie in $\mathcal{A}_F$. Because the uniform conditional distribution on $\{S_{N_j+k} = \gamma_j\}$ is exchangeable (invariant under permutations of the N_j+k

coordinates, as noted in Diaconis–Freedman [9, Section 1]), the conditional probability of E_j^c satisfies

$$\begin{aligned}\mathbb{P}(E_j^c \mid S_{N_j+k} = \gamma_j) &\leq \sum_{i=1}^k \mathbb{P}(X_i \notin \mathcal{A}_F \mid S_{N_j+k} = \gamma_j) \\ &= \frac{k \mathbb{E}[M_j \mid S_{N_j+k} = \gamma_j]}{N_j + k} \leq \frac{k m_j^{\max}}{N_j + k} \xrightarrow{j \rightarrow \infty} 0.\end{aligned}\quad (16)$$

On the event E_j , further condition on the pair (I_{off}, U_j) . Given (I_{off}, U_j) with $I_{\text{off}} \subset \{k+1, \dots, N_j+k\}$ and $|I_{\text{off}}| = m \leq m_j^{\max}$, the remaining $M'_j := N_j + k - m$ random variables $(X_i)_{i \notin I_{\text{off}}}$ (which include $X_1, \dots, X_k$) are conditionally a uniform random walk of length M'_j on $\mathcal{A}_F$ conditioned on their sum being

$$\gamma'_j := \gamma_j - U_j \in M'_j \cdot \mathcal{A}_F,$$

so the conditional law of $(X_1, \dots, X_k)$ given $(S_{N_j+k} = \gamma_j, E_j, I_{\text{off}}, U_j)$ is precisely $Q_{F, \gamma'_j, k}$. Moreover, uniformly over all possible realizations of (I_{off}, U_j) on $\{S_{N_j+k} = \gamma_j\}$, the empirical mean $\mu'_j := \gamma'_j / M'_j$ satisfies

$$\mu'_j = \frac{\gamma_j - U_j}{N_j + k - m} = \frac{\mu_j - U_j / (N_j + k)}{1 - m / (N_j + k)} \xrightarrow{j \rightarrow \infty} \mu_\infty \in \text{relint}(F).$$

Consequently, fixing a compact neighborhood $K \subset \text{relint}(F)$ of μ_∞ , we have $\mu'_j \in K$ for every realization of (I_{off}, U_j) whenever j is sufficiently large. Applying Theorem 3.1 of Diaconis–Freedman [9, Theorem 1.1] on $\mathcal{A}_F$, we obtain, uniformly over all realizations of (I_{off}, U_j) on E_j ,

$$\begin{aligned}&\left| \mathbb{E}[g(S_k) \mid S_{N_j+k} = \gamma_j, E_j, I_{\text{off}}, U_j] - \mathbb{E}_{P_{F, \theta(\mu'_j)}}[g(S_k)] \right| \\ &\leq 2 \|g\|_\infty \left\| Q_{F, \gamma'_j, k} - P_{F, \theta(\mu'_j)}^{\otimes k} \right\|_{\text{TV}}\end{aligned}\quad (17)$$

$$\leq \frac{2 \|g\|_\infty C_K k}{N_j + k - m_j^{\max}} \xrightarrow{j \rightarrow \infty} 0,\quad (18)$$

where $\|g\|_\infty := \max_{b \in k \cdot \mathcal{A}} |g(b)| < \infty$. By the continuity of $\mu' \mapsto \theta(\mu') = \mu_F^{-1}(\mu')$ in Theorem 1.2(ii), $\mathbb{E}_{P_{F, \theta(\mu'_j)}}[g(S_k)] \rightarrow \mathbb{E}_{P_{F, \theta(\mu_\infty)}}[g(S_k)]$ uniformly as

$j \rightarrow \infty$. Combining this with $\mathbb{P}(E_j^c \mid S_{N_j+k} = \gamma_j) \rightarrow 0$ from (16) and the law of total expectation yields

$$\lim_{j \rightarrow \infty} \mathbb{E}[g(S_k) \mid S_{N_j+k} = \gamma_j] = \mathbb{E}_{P_{F, \theta(\mu_\infty)}}[g(S_k)] \geq \delta,$$

which contradicts (14). Finally, since every coefficient of $(\sum_{a \in \mathcal{A}} t^a)^N f(t)$ on $(N+k) \cdot \mathcal{A}$ is strictly positive for $N \geq N_0$, its support equals $(N+k) \cdot \mathcal{A}$ and its Newton polytope equals $\text{conv}((N+k) \cdot \mathcal{A})$. $\square$

References

- [1] L. Baldi, R. Sinn, M. L. Telek, and J. Weigert, *Toric extensions of Pólya's theorem*, arXiv preprint arXiv:2511.07342 (2025).
- [2] O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, John Wiley & Sons, Chichester, 1978.
- [3] R. N. Bhattacharya and R. Ranga Rao, *Normal Approximation and Asymptotic Expansions*, John Wiley & Sons, New York, 1976; reprinted and updated by Robert E. Krieger Publishing Co., Malabar, FL, 1986.
- [4] M. W. Birch, *Maximum likelihood in three-way contingency tables*, Journal of the Royal Statistical Society, Series B (Methodological) **25** (1963), no. 1, 220–233.
- [5] L. D. Brown, *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 9, Institute of Mathematical Statistics, Hayward, CA, 1986.
- [6] M. Castle, V. Powers, and B. Reznick, *Pólya's Theorem with zeros*, Journal of Symbolic Computation **46** (2011), no. 9, 1039–1048.
- [7] I. Csiszár, *Sanov property, generalized I-projection and a conditional limit theorem*, The Annals of Probability **12** (1984), no. 3, 768–793.
- [8] I. Csiszár and F. Matúš, *Generalized maximum likelihood estimates for exponential families*, Probability Theory and Related Fields **141** (2008), 213–246.

- [9] P. Diaconis and D. A. Freedman, *Conditional limit theorems for exponential families and finite versions of de Finetti's theorem*, Journal of Theoretical Probability **1** (1988), no. 4, 381–410.
- [10] P. J. C. Dickinson and J. Povh, *On an extension of Pólya's Positivstellensatz*, Journal of Global Optimization **61** (2015), 615–625.
- [11] F. Esscher, *On the probability function in the collective theory of risk*, Skandinavisk Aktuarietidskrift **15** (1932), 175–195.
- [12] D. Geiger, C. Meek, and B. Sturmfels, *On the toric algebra of graphical models*, The Annals of Statistics **34** (2006), no. 3, 1463–1492.
- [13] C. J. Geyer, *Likelihood inference in exponential families and directions of recession*, Electronic Journal of Statistics **3** (2009), 259–289.
- [14] P. R. Halmos and L. J. Savage, *Application of the Radon-Nikodym Theorem to the theory of sufficient statistics*, The Annals of Mathematical Statistics **20** (1949), no. 2, 225–241.
- [15] D. Handelman, *Representing polynomials by positive linear functions on compact convex polyhedra*, Pacific Journal of Mathematics **132** (1988), no. 1, 35–62.
- [16] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed., Springer Texts in Statistics, Springer, New York, 2005.
- [17] A. Martin-Löf, *Statistical Mechanics and the Foundations of Thermodynamics*, Lecture Notes in Physics, vol. 101, Springer-Verlag, Berlin, 1979.
- [18] V. V. Petrov, *Sums of Independent Random Variables*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 82, Springer-Verlag, Berlin, 1975.
- [19] G. Pólya, *Über positive Darstellung von Polynomen*, Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich **73** (1928), 141–145.
- [20] J. Rauh, T. Kahle, and N. Ay, *Support sets in exponential families and oriented matroid theory*, International Journal of Approximate Reasoning **52** (2011), no. 5, 613–626.

- [21] E. L. Rvaceva, *On domains of attraction of multi-dimensional distributions*, Selected Translations in Mathematical Statistics and Probability **2** (1962), 183–205.
- [22] F. Spitzer, *Principles of Random Walk*, 2nd ed., Graduate Texts in Mathematics, vol. 34, Springer-Verlag, New York, 1976.
- [23] J. M. Van Campenhout and T. M. Cover, *Maximum entropy and conditional probability*, IEEE Transactions on Information Theory **27** (1981), no. 4, 483–489.
- [24] M. J. Wainwright and M. I. Jordan, *Graphical Models, Exponential Families, and Variational Inference*, Foundations and Trends in Machine Learning **1** (2008), no. 1–2, 1–305.
- [25] S. L. Zabell, *A. P. Dempster's "Direct Probabilities" and the Gibbs conditioning principle*, Journal of Statistical Planning and Inference **35** (1993), no. 3, 349–362.