SHORT PROOF OF THE BERNOULLI CONJECTURE

Abstract. We give a short proof of the Bernoulli conjecture. The proof was produced by an internal model at OpenAI.

1. Introduction

Let $\varepsilon = (\varepsilon_i)_{i\in I}$ have independent uniform sign coordinates, and let $G = (g_i)_{i\in I}$ have independent standard Gaussian coordinates. For a nonempty $T \subset \ell^2(I)$, write

$$b(T) = \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i\varepsilon_i, \qquad g(T) = \mathbb{E}\sup_{t\in T}\sum_{i\in I} t_i g_i,$$

with the usual finite-subset interpretation. Recall that $X \preceq_{\mathrm{cx}} Y$ means $\mathbb{E}\Phi(X) \le \mathbb{E}\Phi(Y)$ for every convex Φ for which the expectations are defined.

Theorem 1.1 (Bernoulli conjecture)**.** *There is a universal constant C such that, whenever $\varnothing \neq T \subset \ell^2(I)$ and $b(T) < \infty$, there exist $U \subset \ell^1(I)$ and $V \subset \ell^2(I)$ such that*

$$T \subset U + V, \qquad \sup_{u\in U}\|u\|_1 \le Cb(T), \qquad g(V) \le Cb(T).$$

The conjecture was proposed by Talagrand [9] and first proved by Bednorz and Latała [2]. The main input for our proof is a subgaussian convex-comparison theorem, which follows by duality from Liu's comparison principle [5, Corollary 2] and was explicitly formulated in convex order by van Handel [10]. In fact, the key input for Theorem 1.1 is the following convex domination result.

Theorem 1.2 (Gaussian-to-Bernoulli comparison)**.** *There is a universal constant C_0 such that, for every $G \sim N(0, I_n)$ and every uniform $\varepsilon \in \{-1,1\}^n$,*

$$X \preceq_{\mathrm{cx}} G, \qquad \|X\|_\infty \le 1 \quad \textit{almost surely} \quad \Longrightarrow \quad X \preceq_{\mathrm{cx}} C_0\varepsilon.$$

We use this Gaussian comparison in the following form: for a universal C_G, if Y is centered and

$$\log \mathbb{E}e^{\langle a,Y\rangle} \le \frac{K^2}{2}\|a\|_2^2 \qquad (a \in \mathbb{R}^n), \tag{1}$$

then $Y \preceq_{\mathrm{cx}} C_G K G'$, where $G' \sim N(0, I_n)$ [10]. By Strassen's theorem [8], this is equivalent to the existence of a martingale coupling with

$$Y = C_G K\,\mathbb{E}[G' \mid Y]. \tag{2}$$

Acknowledgment. We thank Ramon van Handel for various helpful comments.

Note on concurrent work: During the preparation of this manuscript, we learned of an independent alternate proof of the Bernoulli conjecture by Liu and Zadik [6].

Date: August 11, 2026.

2020 *Mathematics Subject Classification.* 60G50, 60G15, 60E15, 46B09.

Key words and phrases. Bernoulli theorem, convex comparison, Gaussian processes, randomized rounding.

2. Reduction to Theorem 1.2

We first give the reduction from Theorem 1.1 to Theorem 1.2.

Proof. By the standard finite-section and compactness reduction, it suffices to prove Theorem 1.1 for a finite $T \subset \mathbb{R}^n$. Set

$$D(T) = \inf_{(v_t)_{t\in T}} \left\{ \max_{t\in T} \|t - v_t\|_1 + \mathbb{E} \max_{t\in T} \langle v_t, G\rangle \right\}. \tag{3}$$

For a measurable probability kernel $q = (q_t)_{t\in T}$, put

$$q_t \geq 0, \qquad \sum_{t\in T} q_t = 1, \qquad m_t = \mathbb{E}[G q_t(G)].$$

We claim that

$$D(T) = \sup_q \left\{ \sum_{t\in T} \langle t, m_t\rangle : \sum_{t\in T} \|m_t\|_\infty \leq 1 \right\}. \tag{4}$$

Indeed,

$$\max_t \|t - v_t\|_1 = \sup_{\sum_t \|z_t\|_\infty \leq 1} \sum_t \langle t - v_t, z_t\rangle,$$
$$\mathbb{E} \max_t \langle v_t, G\rangle = \sup_q \sum_t \langle v_t, m_t\rangle.$$

The kernels form a weakly compact convex subset of $L^2(\gamma_n)^T$, so one application of minimax gives

$$D(T) = \sup_{\substack{\sum_t \|z_t\|_\infty \leq 1 \\ q_t \geq 0,\ \sum_t q_t = 1}} \inf_{(v_t)} \left\{ \sum_t \langle t, z_t\rangle + \sum_t \langle v_t, m_t - z_t\rangle \right\}.$$

The inner infimum is finite only when $z_t = m_t$ for every t: if $m_s \neq z_s$, take $v_s = -R(m_s - z_s)$ and $v_t = 0$ for $t \neq s$ and let $R \to \infty$. In the remaining case the inner expression is $\sum_t \langle t, m_t\rangle$. This proves (4) and, in particular, explains precisely why the dual variable is forced to equal the Gaussian barycenter.

Fix an admissible kernel in (4), and write $p_t = \mathbb{E} q_t(G)$ and $a_t = \|m_t\|_\infty$. Since $\sum_t (a_t - p_t)_+ \leq \sum_t a_t \leq 1$, choose a probability vector $r = (r_t)_{t\in T}$ such that $r_t \geq (a_t - p_t)_+$. Toss an independent fair coin. On heads, sample a label J from $q(G)$; on tails, sample J independently of G with distribution r. Then

$$\mathbb{P}(J = t) = \frac{p_t + r_t}{2}, \qquad \mathbb{E}\big[G \mathbf{1}_{\{J=t\}}\big] = \frac{m_t}{2}.$$

Consequently $X = \mathbb{E}[G \mid J]$ satisfies $X \preceq_{\mathrm{cx}} G$, and on every positive-probability event $\{J = t\}$,

$$X = \frac{m_t}{p_t + r_t}, \qquad \|X\|_\infty = \frac{a_t}{p_t + r_t} \leq 1.$$

Apply Theorem 1.2 to the convex function $\Phi(x) = \max_{t\in T} \langle t, x\rangle$. We obtain

$$\frac{1}{2} \sum_{t\in T} \langle t, m_t\rangle = \mathbb{E}\langle J, X\rangle \leq \mathbb{E}\Phi(X) \leq C_0 b(T).$$

Taking the supremum in (4) yields $D(T) \leq 2C_0 b(T)$. The infimum in (3) is attained by coercivity. For a minimizer (v_t) set $U = \{t - v_t : t \in T\}$ and $V = \{v_t : t \in T\}$. Both terms in (3) are nonnegative. Hence

$$T \subset U + V, \qquad \sup_{u\in U} \|u\|_1 \leq 2C_0 b(T), \qquad g(V) \leq 2C_0 b(T),$$

as required. □

3. Proof of Theorem 1.2

The first lemma is the finite-support Gaussian case of a maximum-entropy martingale bridge. It is a special case of the construction of Hua, Song, and Tudose [4, Proposition 3.3 and Section 4.1], applied to $X/2 \preceq_{\mathrm{cx}} G/2$. The general theory of martingale Schrödinger bridges is developed by Nutz and Wiesel [7] and by Backhoff, Beiglböck, Bifronte, and Ley [1]. We include the short argument for completeness.

Lemma 3.1. *Suppose X has finite support and $X \preceq_{\mathrm{cx}} G$. There exists a coupling of X and $G \sim N(0, I_n)$ such that*

$$\mathbb{E}[G \mid X] = \tfrac{1}{2}X, \tag{5}$$

and each conditional distribution $G \mid \{X = x\}$ has an everywhere positive, 1-strongly log-concave density. Among couplings satisfying (5), this coupling uniquely maximizes $H(X \mid G)$.

Proof. Write $\mathbb{P}(X = x_j) = p_j > 0$ for $1 \le j \le m$. Convex comparison gives $\mathbb{E}X = 0$. Set $a_m = 0$, $b_m = 0$, and $\ell_j(g) = a_j + \langle b_j, g\rangle$. Consider

$$P(a,b) = \mathbb{E}\log \sum_{j=1}^{m} e^{\ell_j(G)} - \sum_{j=1}^{m} p_j \left(a_j + \tfrac{1}{2}\langle b_j, x_j\rangle\right). \tag{6}$$

Let $h = \mathbb{E}\max_j \ell_j(G)$. Since log-sum-exp dominates the maximum,

$$P(a,b) \ge \tfrac{1}{2}\left(h - \sum_j p_j \ell_j(x_j)\right) + \tfrac{1}{2}\mathbb{E}\left[\max_j \ell_j(G) - \sum_j p_j \ell_j(G)\right]. \tag{7}$$

The first term is nonnegative by $X \preceq_{\mathrm{cx}} G$. The second is continuous, positively homogeneous, and strictly positive unless all ℓ_j coincide, hence unless $(a,b) = 0$. Compactness of the unit sphere now shows that P is coercive, so it has a minimizer (a^*, b^*).

Define

$$q_j(g) = \frac{e^{a_j^* + \langle b_j^*, g\rangle}}{\sum_{k=1}^{m} e^{a_k^* + \langle b_k^*, g\rangle}}. \tag{8}$$

Differentiating (6) gives

$$\mathbb{E}q_j(G) = p_j, \qquad \mathbb{E}[G q_j(G)] = \tfrac{1}{2}p_j x_j \qquad (1 \le j < m).$$

The identities for $j = m$ follow by summation from $\mathbb{E}G = \mathbb{E}X = 0$. Conditional on $G = g$, assign $X = x_j$ with probability $q_j(g)$. This has the prescribed X marginal and satisfies (5).

To see the maximum-entropy claim, let $\tilde{q}$ be any other kernel with the same marginal and barycenter constraints, and write $Z(g) = \sum_k e^{a_k^* + \langle b_k^*, g\rangle}$ and $H(\tilde{q}) = -\mathbb{E}\sum_j \tilde{q}_j(G)\log \tilde{q}_j(G)$. Since $\log q_j(g) = a_j^* + \langle b_j^*, g\rangle - \log Z(g)$, those constraints give

$$\mathbb{E}\sum_j \tilde{q}_j(G)\log q_j(G) = \sum_j p_j\left(a_j^* + \tfrac{1}{2}\langle b_j^*, x_j\rangle\right) - \mathbb{E}\log Z(G) = \mathbb{E}\sum_j q_j(G)\log q_j(G).$$

Thus the cross-entropy term is the same for every feasible kernel, and

$$H(q) - H(\tilde{q}) = \mathbb{E}\sum_j \tilde{q}_j(G)\log\frac{\tilde{q}_j(G)}{q_j(G)} \ge 0.$$

Equality holds only when $\tilde{q} = q$ almost surely. Finally, $\log q_j$ is concave because log-sum-exp is convex. Hence the conditional Gaussian density $g \mapsto [p_j(2\pi)^{n/2}]^{-1} q_j(g) e^{-\|g\|_2^2/2}$ is everywhere positive and 1-strongly log-concave. □

The following is an immediate consequence of the Brascamp–Lieb variance inequality [3, Theorem 4.1].

Lemma 3.2. *For the coupling in Lemma 3.1,*

$$\mathrm{Var}(G_i \mid X) \le 1, \qquad \mathbb{E}[G_i^2 \mid X] \le 1 + \tfrac{1}{4} X_i^2.$$

We next record a standard subgaussian estimate needed for Lemma 3.4.

Lemma 3.3. *Let $V_1, \dots, V_n$ be independent and centered, with $\mathbb{E}e^{|V_i|/\kappa} \le 2$, and let $Y_i = \mathbb{E}[V_i \mid \mathcal{H}]$ satisfy $|Y_i| \le D$. Then*

$$\log \mathbb{E}e^{\langle a, Y\rangle} \le \max\{4\kappa^2, 2\kappa D\} \|a\|_2^2 \qquad (a \in \mathbb{R}^n).$$

Proof. The standard Bernstein moment estimate for a centered subexponential variable [11, proof of Proposition 2.8.1] gives

$$\mathbb{E}e^{sV_i} \le e^{4\kappa^2 s^2} \qquad \text{if } \kappa|s| \le \tfrac{1}{2}.$$

Set $A = \{i : \kappa|a_i| > 1/2\}$. For $i \in A$, $D|a_i| \le 2\kappa D a_i^2$. Thus conditional Jensen and independence give

$$\mathbb{E}e^{\langle a, Y\rangle} \le e^{2\kappa D \sum_{i \in A} a_i^2} \prod_{i \notin A} \mathbb{E}e^{a_i V_i} \le \exp\left(2\kappa D \sum_{i \in A} a_i^2 + 4\kappa^2 \sum_{i \notin A} a_i^2\right).$$

The claimed bound follows. □

Lemma 3.4. *Suppose $\|X\|_\infty \le 1$, and use the coupling of Lemma 3.1. For every $L > 0$, there exist a uniform $\varepsilon \in \{-1, 1\}^n$ and a centered $Y \in \mathbb{R}^n$ such that*

$$\tfrac{1}{2} X = L\mathbb{E}[\varepsilon \mid X] + Y, \qquad \|Y\|_\infty \le \frac{1}{2L}, \qquad \log \mathbb{E}e^{\langle a, Y\rangle} \le \frac{4}{L^2}\|a\|_2^2. \tag{9}$$

Proof. Independently of (G, X), take independent $U_i \sim \mathrm{Unif}[-L, L]$ and set $\varepsilon_i = \mathrm{sgn}(G_i + U_i)$. The pairs (G_i, U_i) are independent and symmetric. Therefore the ε_i are genuinely independent uniform signs; no independence conditional on X is asserted or needed. Moreover,

$$\mathbb{E}[\varepsilon_i \mid G, X] = \mathrm{clip}(G_i/L, -1, 1).$$

Define $V_i = \mathrm{sgn}(G_i)(|G_i| - L)_+$ and $Y_i = \mathbb{E}[V_i \mid X]$. The V_i are independent and centered, and $G_i = L\,\mathrm{clip}(G_i/L, -1, 1) + V_i$. Taking conditional expectations and using (5) proves the first identity in (9).

The elementary inequality $(r - L)_+ \le r^2/(4L)$ gives

$$\mathbb{E}e^{L|V_i|} \le \mathbb{E}e^{G_i^2/4} = \sqrt{2} \le 2.$$

Lemma 3.2 and $|X_i| \le 1$ further give

$$|Y_i| \le \frac{\mathbb{E}[G_i^2 \mid X]}{4L} \le \frac{1 + X_i^2/4}{4L} \le \frac{1}{2L}.$$

Apply Lemma 3.3 with $\kappa = L^{-1}$ and $D = (2L)^{-1}$ to obtain the final bound in (9). □

Proof of Theorem 1.2. For a convex $\phi : \mathbb{R}^n \to \mathbb{R}$, write

$$S_\phi = \sup\big\{\mathbb{E}\phi(X) : X \preceq_{\mathrm{cx}} G,\ \|X\|_\infty \le 1\big\} < \infty.$$

It suffices to show that $S_\phi \le \mathbb{E}\phi(C_0\varepsilon)$ for every convex ϕ. Conditional expectation on increasingly fine finite partitions of the cube shows that the supremum defining S_ϕ may be taken over finite-valued X.

Set $C_* = \max\{2\sqrt{2}C_G, 1/2\}$, fix $L > 2C_*$, and put $\lambda = C_*/L$ and $C_0 = 2L/(1-2\lambda)$. For each finite-valued admissible X, Lemma 3.4 and Gaussian comparison (2) yield a uniform sign vector ε and $G' \sim N(0, I_n)$ such that

$$\tfrac{1}{2}X = L\mathbb{E}[\varepsilon \mid X] + \lambda Z, \qquad Z = \mathbb{E}[G' \mid X].$$

Jensen gives $Z \preceq_{\mathrm{cx}} G$, and Lemma 3.4 gives

$$\|Z\|_\infty \le \frac{1}{2L\lambda} = \frac{1}{2C_*} \le 1, \qquad X = (1-2\lambda)\mathbb{E}[C_0\varepsilon \mid X] + 2\lambda Z.$$

Thus another application of Jensen gives

$$\mathbb{E}\phi(X) \le (1-2\lambda)\mathbb{E}\phi(C_0\varepsilon) + 2\lambda\mathbb{E}\phi(Z) \le (1-2\lambda)\mathbb{E}\phi(C_0\varepsilon) + 2\lambda S_\phi.$$

Taking the supremum over X and using $2\lambda < 1$ yields $S_\phi \le \mathbb{E}\phi(C_0\varepsilon)$, as claimed. □

References

[1] J. Backhoff, M. Beiglböck, G. Bifronte, and A. Ley, *Bridging classical and martingale Schrödinger bridges*, 2026, arXiv:2604.01299.

[2] W. Bednorz and R. Latała, *On the boundedness of Bernoulli processes*, Ann. of Math. (2) **180** (2014), 1167–1203, doi:10.4007/annals.2014.180.3.8.

[3] H. J. Brascamp and E. H. Lieb, *On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation*, J. Funct. Anal. **22** (1976), 366–389, doi:10.1016/0022-1236(76)90004-5.

[4] D. M. Hua, A. Song, and S. Tudose, *On Talagrand's Convexity Conjecture*, 2026, arXiv:2605.10908.

[5] J. Liu, *Simple and Sharp Generalization Bounds via Lifting*, 2025, arXiv:2508.18682.

[6] J. Liu and I. Zadik, *A Bayesian Proof of the Bernoulli Theorem*, 2026, arXiv:2608.11031.

[7] M. Nutz and J. Wiesel, *On the martingale Schrödinger bridge between two distributions*, to appear in Bernoulli, arXiv:2401.05209.

[8] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439, doi:10.1214/aoms/1177700153.

[9] M. Talagrand, *Constructions of majorizing measures, Bernoulli processes and cotype*, Geom. Funct. Anal. **4** (1994), 660–717, doi:10.1007/BF01896658.

[10] R. van Handel, *On the subgaussian comparison theorem*, 2025, arXiv:2512.18588.

[11] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, second ed., Cambridge University Press, Cambridge, 2026, doi:10.1017/9781009490672.