The Price of Jumpiness: Varentropy, Magnitude-Information Profiles, and Finite-Horizon Floor-Breach Risk in Kelly Allocation

Anatoly Vitold Stankyavichyus

✉ For correspondence: anatolyvitold.stankyavichyus@gmail.com

Abstract

Kelly growth theory ranks allocations by an infinite-horizon average, and standard risk controls temper it with variance, volatility, or drawdown surrogates. These scalar summaries miss a finite-horizon path feature that matters under an absorbing floor: whether adverse outcomes accumulate gradually or concentrate in a few low-frequency events. We separate severity (magnitude) from rarity (physical-measure surprisal) through one new object, the loss-side magnitude-information profile

$$K_{-}(u,t) = \mathbb{P}(X \leq -u, \iota(X) \geq t), \qquad \iota(x) = -\log f(x),$$

and a non-asymptotic set-indexed decomposition: the running-minimum floor-breach probability is bounded by a body-accumulation term on paths avoiding any chosen adverse set plus the probability of hitting it. Since surprisal is not observable without distributional assumptions, downside varentropy serves as an estimable witness for high-surprisal loss states; the resulting distribution-free certificate uses only counts, a magnitude-truncated exponential transform, and estimated varentropy, and its Cantelli component is sharp among all bounds from the first two loss-side information moments. A converse shows the loss-side score law cannot by itself locate adverse mass on the magnitude axis, forcing the primitive object to be two-dimensional. In an oracle-score model, a minimax separation shows that, over a class constrained only by loss-side score summaries, the aggressive negative-power moment a strong drawdown certificate requires is minimax-inestimable, while every input of our certificate is estimable at the parametric rate. A controlled matched-pair experiment instantiates the mechanism on synthetic laws matched in mean, variance, and a weak negative-power moment but differing in profile and floor-breach risk.

Keywords: heavy-tailed distributions, kelly criterion, minimax estimation, portfolio optimization, risk management, self-information, varentropy.

Contents

I Introduction

This article is the third in a sequence whose broader goal is to turn varentropy into a practical tool for risk management [1], [2]. The earlier papers in the sequence addressed more structural questions, such as lower bounds for varentropy and the computation of varentropy for non-extremal stable laws. The present paper turns to a more applied question: how can varentropy, and information-theoretic reasoning more generally, help practitioners choose portfolios subject to a finite-horizon floor-breach mandate?

Market participants rarely evaluate portfolios only by the absolute growth rate that might be achieved over a lifetime of investment. They also operate under risk mandates. This is especially important for institutional investors with explicit capital or solvency constraints, such as pension funds, endowments, and insurance companies investing float, and for leveraged hedge funds. One severe log return can make recovery impossible when the investment process contains an absorbing state, such as a margin call, forced deleveraging, or a redemption constraint. Without that absorbing state, the investor might have been able to continue with the same aggressive strategy and eventually recover; with it, the path itself matters. This is why portfolio rules that look only at terminal wealth quantiles or expected terminal wealth can give an incomplete account of risk. In practice, wealth is path dependent.

The central question of this paper is how to formalize the effect of severe rare events on the log-wealth path and on finite-horizon floor-breach risk. One possible answer is to use the existing machinery of power laws and fat-tailed distributions. That answer is often insufficient for the present purpose, because it can miss important pre-asymptotic effects generated by distributions that are not technically fat-tailed, such as mixtures and tempered laws. These pre-asymptotic features may wash out in an infinite-horizon central-limit regime, or under an appropriate functional limit theorem, but real investors face finite horizons and finite floors. For an investor who has already suffered a large forced loss, it is cold comfort that the loss did not come from a "true" asymptotic fat tail.

The approach taken here is to separate two ideas that are often bundled together in fat-tail language: rarity and severity. Rarity has a natural information-theoretic formalization through surprisal, or self-information. Severity is a magnitude or geometry question: log-concavity, s-concavity, tail monotonicity, regular variation, bounded support with a thin edge, and mixture geometry all describe different ways in which adverse magnitudes can be organized. Section III develops a set-indexed framework that can accommodate these geometries without committing to any one of them as the primitive model.

Floor-breach risk and severe rare events are closely connected. Under a "non-jumpy" return law, losses tend to accumulate gradually. Under a jumpier law, one adverse episode may be enough to end the game, so the risk of a floor breach is present from the beginning of the horizon. Equivalently, if gains and losses are concentrated in a small number of events, an investor may not get a second chance. This intuition is related to the single-big-jump, or catastrophe, principle for subexponential distributions, which is recalled in the appendix. The broader message does not require an asymptotic heavy-tail model: all else equal, an investor with a strict risk mandate should care whether a portfolio's profit and loss is concentrated in a few adverse states.

This makes the classical Kelly formulation incomplete for such an investor. Fractional Kelly, variance control, volatility targeting, and related dispersion controls are natural first responses. A more sophisticated approach is the Busseti–Ryu–Boyd risk-constrained Kelly framework [3], which imposes a drawdown-probability constraint through a tractable surrogate. These methods control the level of risk, but they do not by themselves separate finite-horizon paths that reach the floor through a diffuse accumulation of many ordinary losses from paths dominated by a small number of adverse events.

We now formalize the path object. Let $X_1, X_2, \dots$ be one-period log returns, $S_k = \sum_{i=1}^{k} X_i$, and

$$\underline{S}_T = \min_{0\le k\le T} S_k.$$

The event studied in this paper is

$$\mathbb{P}(\underline{S}_T \le -c), \tag{1}$$

for a finite horizon T and floor $c > 0$. Equivalently, this is the finite-time first-passage event $\{\tau_{-c} \le T\}$. It is a *running-minimum floor-breach* event measured relative to initial log wealth. It is not the classical peak-to-trough drawdown

$$\max_{0\le j\le k\le T}(S_j - S_k),$$

although the two are economically related. We reserve the word "drawdown" for the Busseti–Ryu–Boyd (BRB) literature; the theorem itself certifies finite-time floor breach. The running minimum is used because it is the cleanest path functional that records whether the investor survives the bad part of the path. Survival probabilities, running-minimum VaR, forced exit, and Kelly exposure constraints all require bounds on the same event.

There is an important boundary condition. In the finite-variance central-limit regime, with c on the $\sqrt{T}$ scale and $T \to \infty$, Donsker universality [4] makes the running minimum depend only on variance. Shape is then washed out. The effect studied here lives elsewhere: at finite horizons before that universality is reliable, and at large thresholds where the breach is a finite-threshold or large-deviation event. This is also the regime in which bounded but mixture-like laws can be economically jumpy even though they are not asymptotically heavy-tailed.

The organizing theorem is set-indexed. Let $A \subseteq (-\infty, 0)$ be an adverse set of one-period losses. Define

$$B_T(c; A) = \mathbb{P}\{\underline{S}_T \le -c,\ X_i \notin A \text{ for all } i \le T\},$$

the probability of breaching the floor while avoiding A, and

$$H_T(A) = \mathbb{P}\{\exists i \le T : X_i \in A\}.$$

Then

$$\mathbb{P}(\underline{S}_T \le -c) \le B_T(c; A) + H_T(A).$$

Thus finite-horizon path concentration is decomposed into two mechanisms:

$$\text{body accumulation while avoiding } A \qquad + \qquad \text{adverse-set hit risk.}$$

The adverse set may be a magnitude tail $A_u = (-\infty, -u]$, a loss-side information set, or a joint magnitude-information set. The inequality itself is an elementary union bound; its value is organizational. It gives a common language for separating ordinary body accumulation from adverse-event exposure without assuming regular variation, symmetry, or a Lévy model.

The central information-theoretic object is the loss-side magnitude-information profile. For a density f, write

$$\imath(x) = -\log f(x)$$

and define

$$K_{-}(u,t) = \mathbb{P}(X \le -u,\ \iota(X) \ge t).$$

This profile records both the size of a loss and how low-density, or surprising, that loss is under the physical increment law. It is the main conceptual object of the paper. Classical fat-tail analysis studies rare large magnitudes through $\bar{F}_{-}(u)$, or through asymptotic regular variation of that tail. The profile $K_{-}(u,t)$ asks a finer finite-horizon question: which losses are both economically adverse and low-density under the physical measure? Its projections include the ordinary loss tail $\bar{F}_{-}(u)$ and the one-dimensional loss-side information profile $K_{-}(t)$. In monotone loss-side geometries, magnitude tails are recovered as profile slices. In asymmetric, bounded, tempered, or mixture-like laws, the profile can also reveal rare adverse components that a tail index or volatility summary does not isolate.

In real applications, however, the physical law is not known exactly. The surprisal $\iota(x)$ should therefore not be treated as directly observable unless one is willing to impose a distributional model that will almost never be exactly true. This is where varentropy enters. Varentropy is the variance of self-information,

$$\mathrm{VE}(X) = \mathrm{Var}(-\log f(X)),$$

and downside varentropy is the corresponding dispersion of self-information on the loss side. It serves here as an estimable witness for the high-surprisal part of the profile that cannot be observed directly without knowing the true physical density. If high-surprisal loss states have nontrivial probability, the loss-side self-information score must have nontrivial dispersion; conversely, Cantelli's inequality gives the sharp distribution-free upper bound for such high-surprisal loss hits using only that dispersion. It is worth emphasizing that downside varentropy is not meant as a substitute for $K_{-}(u,t)$ but instead it is a conservative scalar certificate used when direct profile estimation would require unjustified distributional assumptions.

This theory has practical consequences for portfolio construction. A Kelly or Busseti–Ryu–Boyd-style allocation can satisfy a drawdown or volatility surrogate while still concentrating its remaining risk in adverse cells. The path-concentration constraint refines the risk mandate by requiring a separation between ordinary body accumulation and adverse-hit exposure. The controlled experiment in Section IV-D isolates this mechanism: our Path Concentration constraint (PC) is compared with variance and Busseti–Ryu–Boyd (BRB) moments on a pair of synthetic laws whose population summaries are matched but whose magnitude-information profiles and floor-breach risks differ.

The paper is organized accordingly. Section II records the minimal self-information identities needed for the score certificate. Section III proves the set-indexed finite-threshold decomposition, the profile calculus, the sharp distribution-free profile and hit-budget envelopes, the profile-to-floor-breach transfer theorem, and the optional Pareto-score refinement. Section IV turns the certificate into survival, risk-budget, Kelly/BRB, and multi-asset allocation constraints and gives the controlled synthetic mechanism experiment. The final sections calibrate score-tail and tail-regular special cases.

II Notations, self-information, and downside varentropy

The practical certificate below uses downside varentropy because it is an integrated self-information functional, not a pointwise tail or density value. For a density f, write

$$\iota(x) = -\log f(x), \qquad h = \mathbb{E}[\iota(X)], \qquad \mathrm{VE}(X) = \mathrm{Var}(\iota(X)).$$

The same definitions are used on the loss side:

$$p_{-} := \mathbb{P}(X < 0), \qquad h_{-} := \mathbb{E}[\iota(X) \mid X < 0], \qquad V_{-} := \mathrm{Var}(\iota(X) \mid X < 0).$$

The scalar V_{-} is the varentropy of the conditional loss-side law, because if f_{-} is the density of $X \mid X < 0$, then $-\log f(x) = -\log f_{-}(x) - \log p_{-}$ on the loss side, and the additive constant has no effect on variance. When $p_{-} = 0$, there is no loss-side conditional law; all loss-side statements below are then vacuous and the path floor cannot be crossed by negative increments. Unless stated otherwise we therefore work on the nontrivial case $p_{-} > 0$.

The two facts needed later are simple. First, full varentropy is scale-free, and thresholded or loss-side varentropy is invariant only when the conditioning region is transformed along with the random variable.

Proposition II.1 (Affine action and threshold-set invariance). *Let $Y = aX + b$, $a \neq 0$, and let f_Y be its density. Then*

$$-\log f_Y(Y) = -\log f_X(X) + \log |a|. \qquad (2)$$

Consequently

$$h(Y) = h(X) + \log|a|, \qquad \mathrm{VE}(Y) = \mathrm{VE}(X).$$

More generally, if $C \subset \mathbb{R}$ is a Borel set with $\mathbb{P}(X \in C) > 0$ and $C_Y = aC + b$, then

$$\mathrm{Var}\big(-\log f_Y(Y) \mid Y \in C_Y\big) = \mathrm{Var}\big(-\log f_X(X) \mid X \in C\big). \tag{3}$$

In particular, downside varentropy relative to a threshold τ,

$$V_{\le\tau}(X) := \mathrm{Var}(\iota_X(X) \mid X \le \tau),$$

is invariant under a positive affine rescaling when the threshold is moved to $a\tau + b$. The conventional notation V_-, which uses the fixed event $X < 0$, is therefore invariant under transformations that preserve or consistently map that loss event; it is not invariant under an arbitrary translation while the threshold zero is held fixed.

Proof. Since $f_Y(y) = |a|^{-1} f_X((y-b)/a)$, substituting $y = Y$ gives (2). The entropy identity follows by taking expectations; the full varentropy identity follows because adding a constant does not change variance. On the event $Y \in C_Y$ we have exactly $X \in C$, and the same constant-shift argument gives (3). □

III Finite-threshold path concentration and score certificates

This section contains the core theorems in the article. Several of the inequalities used below are standard—union bounds, Chernoff/Bernstein bounds, Cantelli's inequality, and Poisson approximation. The novelty is not the individual inequalities themselves, but the way the magnitude-information profile $K_-(u,t)$ turns them into a finite-horizon risk-composition certificate separating body accumulation from adverse-hit exposure. It first splits the finite-horizon breach event into two mechanisms: *body accumulation*, where the path crosses the floor while avoiding a designated adverse region of the loss side, and an *adverse-set hit*, where at least one increment lands in that region. The adverse region need not be a magnitude tail. It may be a loss-side information exceedance, a magnitude threshold, or a joint magnitude-information set. The second half of the section then shows how to certify the adverse-hit term using only statistically estimable loss-side information moments and, if desired, a one-dimensional Pareto-score envelope.

Throughout this section $X_1, X_2, \ldots$ are i.i.d. real increments, $S_k = \sum_{i=1}^k X_i$, $S_0 = 0$, and $\underline{S}_T = \min_{0\le k\le T} S_k$. An *adverse set* is a measurable set $A \subseteq (-\infty, 0)$. Write

$$p_A := \mathbb{P}(X \in A), \qquad D_T(A) := \{\exists i \le T : X_i \in A\}, \qquad H_T(A) := \mathbb{P}(D_T(A)) = 1 - (1-p_A)^T.$$

The magnitude-tail choice is $A_u = (-\infty, -u]$, for which $p_{A_u} = \bar{F}_-(u)$. The information-theoretic choices are introduced in Section III-D.

III-A Adverse-set admissibility

The exact event split below is valid for every measurable adverse set. The explicit body bounds require additional properties of the chosen set.

Definition III.1 (Admissible adverse sets)**.** Let $A \subseteq (-\infty, 0)$ be measurable.

(i) A is an *adverse set* if no regularity is imposed beyond measurability. Such a set is enough for the exact body–hit decomposition.

(ii) A is *Chernoff-admissible at $\theta > 0$* if

$$m_A(\theta) := \mathbb{E}\big[e^{-\theta X}\mathbf{1}_{\{X\notin A\}}\big] < \infty.$$

A Chernoff-admissible set gives a Chernoff bound for the body term on paths that avoid A.

(iii) A is *loss-complete at level $u \ge 0$* if

$$(-\infty, -u] \subseteq A \qquad \text{up to null sets.}$$

Then, after excluding the adverse set A, losses are bounded by u: $X > -u$ almost surely. Loss-complete sets are automatically Chernoff-admissible at every $\theta \ge 0$, since

$$m_A(\theta) \le e^{\theta u}.$$

If X has finite second moment, they also admit Bernstein-type body bounds.

Remark III.2 (Importance of assumptions). Set-indexing is the right conceptual level, but not every set gives the same computational tools. A pure information set $A_t^\iota = \{x < 0 : \iota(x) \geq t\}$ can identify a low-density adverse region, but it need not remove all sufficiently large losses unless the loss-side geometry is monotone. In that case the exact decomposition is still valid, while the Chernoff or Bernstein body bounds require a separate admissibility check. Magnitude tails A_u are loss-complete. Joint information sets $A_{u,t}^\iota$ are loss-complete only in the special case where the information threshold does not remove any part of the magnitude tail; for example, when $t \leq I_-(u)$, so $A_{u,t}^\iota$ coincides with A_u up to null sets.

III-B A set-indexed body–adverse-hit decomposition

Theorem III.3 (Finite-threshold decomposition: set-indexed upper bound)**.** *Let $A \subseteq (-\infty, 0)$ be measurable. For every horizon $T \in \mathbb{N}$ and floor $c > 0$, define*

$$B_T(c; A) := \mathbb{P}\big(\underline{S}_T \leq -c,\ X_i \notin A \text{ for all } 1 \leq i \leq T\big).$$

Then

$$\mathbb{P}(\underline{S}_T \leq -c) \leq B_T(c; A) + H_T(A) \leq B_T(c; A) + T p_A. \tag{4}$$

If $m_A(\theta) < \infty$ for some $\theta \geq 0$, then

$$B_T(c; A) \leq e^{-\theta c} \sum_{k=1}^{T} m_A(\theta)^k. \tag{5}$$

Consequently, with the Chernoff body bound for paths avoiding A

$$\widehat{B}_T(c; A) := \inf_{\theta \geq 0:\, m_A(\theta) < \infty} e^{-\theta c} \sum_{k=1}^{T} m_A(\theta)^k, \tag{6}$$

one has

$$\mathbb{P}(\underline{S}_T \leq -c) \leq \widehat{B}_T(c; A) + T p_A. \tag{7}$$

If A is loss-complete at level u, then $m_A(\theta) \leq e^{\theta u}$ for every $\theta \geq 0$, so $\widehat{B}_T(c; A)$ is finite without any moment-generating-function assumption.

Proof. The breach event decomposes according to whether the path ever hits the adverse set:

$$\{\underline{S}_T \leq -c\} \subseteq \big(\{\underline{S}_T \leq -c\} \cap D_T(A)^c\big) \cup D_T(A).$$

This gives the first inequality in (4). Independence gives $\mathbb{P}(D_T(A)) = 1 - (1 - p_A)^T = H_T(A) \leq T p_A$, giving the second.

For the body bound, if the path breaches by time T without hitting A, then for some $1 \leq k \leq T$, $S_k \leq -c$ and $X_i \notin A$ for all $1 \leq i \leq k$. Thus

$$B_T(c; A) \leq \sum_{k=1}^{T} \mathbb{P}(S_k \leq -c,\ X_1 \notin A, \dots, X_k \notin A).$$

For each summand, Markov's inequality applied to $e^{-\theta S_k}$ gives

$$\begin{aligned}\mathbb{P}(S_k \leq -c,\ X_1 \notin A, \dots, X_k \notin A) &\leq e^{-\theta c} \mathbb{E}\left[e^{-\theta S_k} \prod_{i=1}^{k} \mathbf{1}_{\{X_i \notin A\}}\right] \\ &= e^{-\theta c} m_A(\theta)^k,\end{aligned}$$

where the equality uses independence. Summing over k proves (5); optimizing proves (7). If A is loss-complete at level u, then on $\{X \notin A\}$, $X > -u$, so $e^{-\theta X} \leq e^{\theta u}$. □

Remark III.4 (Adverse-set exclusion versus conditioning). The transform m_A is not the moment generating function of a normalized conditional law. It is an unnormalized transform of the original increment with the indicator $\mathbf{1}_{\{X \notin A\}}$ attached. Thus $m_A(\theta)^k$ contains both the transform of the non-adverse body and the probability of avoiding A for k periods. The exact body term is $B_T(c; A)$, while $\widehat{B}_T(c; A)$ is a computable Chernoff upper bound.

Remark III.5 (Why set-indexing is the right level of abstraction). The magnitude-only split uses $A_u = (-\infty, -u]$. That is still the right choice when the only question is whether one increment is large enough by itself to threaten the floor. But mixtures and asymmetric laws can be jumpy for a different reason: the important adverse region may be a low-frequency, low-density part of the loss side, not simply the farthest left tail. The set-indexed theorem lets the adverse region be chosen as a loss-side information exceedance or a joint magnitude-information set. The same body-versus-hit decomposition then applies without changing the probabilistic argument.

The Chernoff body bound is the most general explicit upper bound for the body term. In finite-variance applications, loss-complete adverse sets also give a recognizable Bernstein bound, because losses are bounded after the adverse set has been excluded.

Corollary III.6 (Bernstein bound for body accumulation with loss-complete adverse sets)**.** *Assume X has finite second moment and let A be loss-complete at level u. Define*

$$Y_A := -X\mathbf{1}_{\{X\notin A\}}, \qquad a_A := \mathbb{E}Y_A, \qquad v_A := \mathrm{Var}(Y_A), \qquad r_A := u - a_A.$$

Then $Y_A \le u$ almost surely and $Y_A - a_A \le r_A$. With the convention that the summand below is 1 when $(c - ka_A)_+ = 0$,

$$B_T(c;A) \le \widetilde{B}_T(c;A,u) := \sum_{k=1}^{T} \exp\left\{ -\frac{(c-ka_A)_+^2}{2kv_A + \frac{2}{3}r_A(c-ka_A)_+} \right\}. \tag{8}$$

Consequently,

$$\mathbb{P}(\underline{S}_T \le -c) \le \min\{\widehat{B}_T(c;A), \widetilde{B}_T(c;A,u)\} + Tp_A. \tag{9}$$

If $a_A \le 0$, for instance when $\mathbb{E}X \ge 0$ and A removes only losses, then

$$B_T(c;A) \le T\exp\left\{ -\frac{c^2}{2Tv_A + \frac{2}{3}r_A c} \right\}. \tag{10}$$

For $A = A_u = (-\infty, -u]$, this recovers the magnitude-tail Bernstein body bound.

Proof. If the path breaches at time k while avoiding A, then $X_i \notin A$ for $1 \le i \le k$ and $S_k \le -c$. Hence $\sum_{i=1}^k Y_{A,i} \ge c$, where $Y_{A,i}$ are independent copies of Y_A. Therefore

$$B_T(c;A) \le \sum_{k=1}^{T} \mathbb{P}\left(\sum_{i=1}^{k} Y_{A,i} \ge c \right).$$

For each fixed k, apply the one-sided Bernstein inequality to the centered variables $Y_{A,i} - a_A$, which have total variance kv_A and are bounded above by r_A. This gives (8). Substitution into (4) gives (9). If $a_A \le 0$, then $(c - ka_A)_+ \ge c$. The Bernstein exponent is increasing in the threshold and decreasing in k, so every summand is bounded by the $k = T$, threshold c expression in (10). □

Remark III.7 (Why this is only a body bound)*.* Corollary III.6 is not a tail bound for the original walk. It is a bound for the body mechanism after increments in the adverse set have been excluded. The adverse-hit exposure is bounded by $H_T(A)$, or by Tp_A in the linearized bound. This is why the theorem compares body accumulation against adverse-set exposure rather than trying to describe the whole path with one concentration inequality.

III-C Regime diagnostics and timing

The exact decomposition already gives a useful risk-composition diagnostic. If

$$\mathcal{R}_T^{\mathrm{hit}}(c;A) := \frac{H_T(A)}{B_T(c;A)}$$

is large, the conservative certificate assigns substantial weight to adverse-set exposure relative to body accumulation along paths that avoid the adverse set. If this ratio is small, the body-accumulation term is the main bounded contributor. In applications B_T is unknown, so we use the computable diagnostic

$$\widehat{\mathcal{R}}_T(c;A) := \frac{Tp_A}{\widehat{B}_T(c;A)}.$$

This is a rare-hit, upper-bound diagnostic. A large hatted ratio is evidence that adverse-set exposure is material in the certificate. A small hatted ratio is not proof of body dominance, because Tp_A may overstate the hit term and $\widehat{B}_T$ may overstate the body term.

There is also a timing issue. The event $D_T(A)$ only says that the path hits A sometime before the horizon. It may include paths that breach first and hit A only later. Therefore $H_T(A)$ is a conservative bound for adverse exposure, not by itself a causal attribution of the breach.

For causal attribution define the stopping times

$$\tau_c := \inf\{k \geq 0 : S_k \leq -c\}, \qquad \tau_A := \inf\{k \geq 1 : X_k \in A\},$$

with the usual convention that an empty infimum is ∞. Then

$$\mathbb{P}(\tau_c \leq T) = B_T^{\preceq}(c; A) + J_T(c; A), \tag{11}$$

where

$$B_T^{\preceq}(c; A) := \mathbb{P}(\tau_c \leq T,\ \tau_A > \tau_c), \qquad J_T(c; A) := \mathbb{P}(\tau_c \leq T,\ \tau_A \leq \tau_c).$$

The term $J_T(c; A)$ is the probability of a breach path on which the adverse set is hit before or at the first floor crossing. It satisfies

$$J_T(c; A) \leq H_T(A),$$

and the same Chernoff proof gives

$$B_T^{\preceq}(c; A) \leq \widehat{B}_T(c; A).$$

Thus the master upper bound remains valid with a causal interpretation: the body bound controls paths that cross before any adverse hit, while $H_T(A)$ upper bounds the pre-breach adverse-hit term. For ex post attribution, the relevant conditional share is

$$\mathbb{P}(\tau_A \leq \tau_c \mid \tau_c \leq T), \tag{12}$$

not $\mathbb{P}(D_T(A) \mid \tau_c \leq T)$, which also counts post-breach hits.

This is the general finite-threshold version of the single-big-jump diagnostic: in a heavy-tail calibration A is a magnitude tail; in a mixture law it may be a rare loss-side component; in a score-based implementation it is a high-information loss cell.

III-D The loss-side magnitude-information profile

The set-indexed theorem becomes information-theoretic once adverse sets are chosen from the self-information coordinate. Let X have density f and define $\imath(x) = -\log f(x)$. The primary object is the joint loss-side magnitude-information profile

$$K_-(u, t) := \mathbb{P}(X \leq -u,\ \imath(X) \geq t), \qquad u \geq 0, \tag{13}$$

which records both adverse size and loss-side surprise. Its one-dimensional information marginal is

$$K_-(t) := K_-(0, t) = \mathbb{P}(X < 0,\ \imath(X) \geq t), \tag{14}$$

and the ordinary magnitude tail is the other projection $\bar{F}_-(u) = \mathbb{P}(X \leq -u)$.

Define the canonical information adverse sets

$$A_t^{\imath} := \{x < 0 : \imath(x) \geq t\}, \qquad A_{u,t}^{\imath} := \{x \leq -u : \imath(x) \geq t\}. \tag{15}$$

Then $\mathbb{P}(X \in A_t^{\imath}) = K_-(t)$ and $\mathbb{P}(X \in A_{u,t}^{\imath}) = K_-(u, t)$. The set-indexed decomposition gives

$$\mathbb{P}(\underline{S}_T \leq -c) \leq B_T(c; A_t^{\imath}) + T K_-(t), \tag{16}$$

$$\mathbb{P}(\underline{S}_T \leq -c) \leq B_T(c; A_{u,t}^{\imath}) + T K_-(u, t). \tag{17}$$

The corresponding exact linearized information ratios are

$$\mathcal{R}_T^{\imath}(c, t) := \frac{T K_-(t)}{B_T(c; A_t^{\imath})}, \tag{18}$$

$$\mathcal{R}_T^{\mathrm{joint}}(c, u, t) := \frac{T K_-(u, t)}{B_T(c; A_{u,t}^{\imath})}. \tag{19}$$

The breach-capable part of a joint information set is also expressed through K_-:

$$q_{A_{u,t}^{\imath}}(c, b) = K_-\big(\max\{u, c + b\}, t\big). \tag{20}$$

The magnitude tail is contained in the same profile. Let

$$I_-(u) := \operatorname*{ess\,inf}_{x \leq -u} \imath(x). \tag{21}$$

Then, up to null sets,

$$\bar{F}_{-}(u) = K_{-}(u, I_{-}(u)). \tag{22}$$

Moreover

$$\bar{F}_{-}(u) \le K_{-}(I_{-}(u)). \tag{23}$$

If the loss-side density is strictly monotone as one moves left, then $I_{-}(u) = \imath(-u)$ and the one-dimensional identity sharpens to

$$\bar{F}_{-}(u) = K_{-}(\imath(-u)). \tag{24}$$

Theorem III.8 (Profile calculus for canonical rare-event geometries)**.** *Let $L = -X$ denote loss magnitude, so that the unconditional loss-side density is $g(\ell) = f(-\ell)$ for $\ell > 0$. The magnitude-information profile has the following canonical reductions.*

(i) Power-law loss tails. *Suppose that, for $\ell \ge x_m$,*

$$g(\ell) = A\ell^{-(\beta+1)}, \qquad A > 0, \quad \beta > 0.$$

Then $\imath(-\ell) = -\log A + (\beta + 1)\log \ell$. Hence, for $u \ge x_m$ and for score thresholds t such that

$$x(t) := \exp\left(\frac{t + \log A}{\beta + 1}\right) \ge x_m,$$

one has the exact identity

$$K_{-}(u, t) = \bar{F}_{-}\big(\max\{u, x(t)\}\big). \tag{25}$$

More generally, if $g(\ell) = \ell^{-(\beta+1)+o(1)}$ and the loss-side score is eventually increasing, then the score-implied threshold $x(t)$, defined by $\imath(-x(t)) = t$, satisfies

$$\log x(t) = \frac{t}{\beta + 1} + o(t),$$

and $K_{-}(u, t)$ is the ordinary loss tail evaluated at the larger of the magnitude threshold and the score-implied threshold, up to the corresponding regular-variation error.

(ii) Bounded polynomial edges. *Suppose $L \le M$ and, as $\ell \uparrow M$,*

$$g(\ell) = C(M - \ell)^q(1 + o(1)), \qquad C > 0, \quad q > 0.$$

Let

$$r_t := C^{-1/q}e^{-t/q}.$$

Then, uniformly for fixed $u < M$,

$$K_{-}(u, t) = \int_{\max\{u, M - r_t(1+o(1))\}}^{M} g(\ell)\,\mathrm{d}\ell. \tag{26}$$

In particular, for every fixed $u < M$, as $t \to \infty$,

$$K_{-}(u, t) \sim \frac{C^{-1/q}}{q + 1}\, e^{-(q+1)t/q}. \tag{27}$$

Thus bounded losses can have an exponentially thin score tail even though no asymptotic magnitude tail exists beyond the endpoint.

(iii) Disjoint rare loss components. *Suppose the density contains a disjoint loss-side component,*

$$f(x) = (1 - \varepsilon)g_0(x) + \varepsilon h(x), \qquad 0 < \varepsilon < 1,$$

where the support interval $J \subset (-\infty, 0)$ of h is disjoint from the support of g_0, and suppose $h(x) \le M_h$ on J. Then, for every $u \ge 0$ and every $t \le -\log(\varepsilon M_h)$,

$$K_{-}(u, t) \ge \varepsilon \int_{J \cap (-\infty, -u]} h(x)\,\mathrm{d}x. \tag{28}$$

Consequently a low-frequency adverse component contributes at score level of order $\log(1/\varepsilon)$, even when its magnitudes are bounded or when no asymptotic tail index is meaningful.

Proof. For (i), the exact power-law density gives $\iota(-\ell) = -\log A + (\beta+1)\log \ell$. The joint event $\{X \le -u, \iota(X) \ge t\}$ is therefore the same as $\{L \ge u, L \ge x(t)\}$, which proves (25). In the regularly varying extension, $\iota(-\ell) = (\beta+1)\log \ell + o(\log \ell)$. Eventual monotonicity lets us invert the score coordinate; the inverse satisfies $\log x(t) = t/(\beta+1) + o(t)$, and the same set identity holds with the regular-variation error inherited from the tail approximation.

For (ii), write $r = M - \ell$. Near the endpoint,

$$\iota(-\ell) = -\log C - q\log r + o(1).$$

Thus $\iota(-\ell) \ge t$ is equivalent to $r \le C^{-1/q} e^{-t/q}(1+o(1)) = r_t(1+o(1))$. Intersecting this score event with $\ell \ge u$ gives (26). If $u < M$ is fixed, then eventually $r_t < M - u$, and

$$K\_(u,t) \sim \int_0^{r_t} C r^q \, \mathrm{d}r = \frac{C}{q+1} r_t^{q+1} = \frac{C^{-1/q}}{q+1} e^{-(q+1)t/q},$$

which is (27).

For (iii), on J disjointness gives $f(x) = \varepsilon h(x) \le \varepsilon M_h$, so $\iota(x) = -\log f(x) \ge -\log(\varepsilon M_h)$. Hence, for $t \le -\log(\varepsilon M_h)$, every point of $J \cap (-\infty, -u]$ belongs to the event defining $K\_(u,t)$. Integrating $f = \varepsilon h$ over that set gives (28). □

III-E Distribution-free profile bounds from estimable information moments

While the profile $K\_(u,t)$ is the exact object, a practical theory should not make its basic certificate depend on pointwise density levels such as $\iota(-u)$ or on lower support values such as $\operatorname{ess\,inf}_{X<0} \iota(X)$. Those are population quantities, but they are not robust, distribution-free statistical targets from a raw finite sample. The quantities that motivated the use of entropy and varentropy are integrated log-density functionals: they can be estimated by standard spacing methods [5] or nearest-neighbor methods [6], or from an out-of-sample predictive density, without estimating a tail index or a pointwise density at a selected loss threshold.

Accordingly, the distribution-free practical bound in this paper uses only the following loss-side information moments and ordinary tail/body quantities:

$$p_- := \mathbb{P}(X < 0), \qquad h_- := \mathbb{E}[\iota(X) \mid X < 0], \qquad V_- := \operatorname{Var}(\iota(X) \mid X < 0), \tag{29}$$

plus the empirical loss tail $\bar{F}\_(u) = \mathbb{P}(X \le -u)$ and the magnitude-truncated exponential transform

$$m_u(\theta) := \mathbb{E}\big[e^{-\theta X} \mathbf{1}_{\{X > -u\}}\big]. \tag{30}$$

The mass p_- and tail $\bar{F}\_(u)$ are estimated by counts; the transform $m_u(\theta)$ is an ordinary sample average; and h_- and V_- are loss-side entropy and varentropy functionals. If f_- denotes the conditional loss-side density, then for $x < 0$, $f(x) = p_- f_-(x)$, so

$$-\log f(x) = -\log f_-(x) - \log p_-.$$

Thus V_- is exactly the varentropy of the conditional loss-side law; the additive $-\log p_-$ changes entropy but not varentropy.

For $a > 0$, define the loss-side high-information event

$$A_a^- := \{X < 0 : \ \iota(X) - h_- \ge a\}. \tag{31}$$

This event uses an excess self-information threshold rather than a pointwise threshold $\iota(-u)$. Its probability is controlled by downside varentropy in the sharp distribution-free way below.

Proposition III.9 (Varentropy as a witness for high-surprisal loss states)**.** *Let $Y = \iota(X) \mid \{X < 0\}$ have mean h_- and variance V_-. If, for some $a > 0$,*

$$q_a := \mathbb{P}\{Y - h_- \ge a\} > 0,$$

then

$$V_- \ge \frac{q_a}{1-q_a}\, a^2. \tag{32}$$

Equivalently, a non-negligible high-surprisal loss-side cell forces downside varentropy to be large enough to pay for it. The converse use of the same algebra is the Cantelli hit certificate below.

Proof. The claim is the contrapositive algebraic form of Cantelli's inequality. For a real random variable with mean h_- and variance V_-, $q_a \le V_-/(V_- + a^2)$. Solving this inequality for V_- gives (32).

□

Theorem III.10 (Exactness of the varentropy witness in a two-shelf loss model)**.** *Let* $Y = \iota(X) \mid \{X < 0\}$ *take two values* $y_0 < y_1$*, with probabilities* $1 - q$ *and* q*, respectively, where* $0 < q < 1$*. Let*

$$h_- = (1-q)y_0 + qy_1, \qquad V_- = q(1-q)(y_1 - y_0)^2,$$

and set

$$a = (1-q)(y_1 - y_0).$$

Then the high-surprisal shelf is exactly the event $\{Y - h_- \ge a\}$*, and*

$$\mathbb{P}\{Y - h_- \ge a\} = q = \frac{V_-}{V_- + a^2}. \tag{33}$$

Equivalently, in the unconditional loss-side event,

$$\mathbb{P}(A_a^-) = p_- q = p_- \frac{V_-}{V_- + a^2}.$$

Thus the Cantelli-varentropy certificate is exact for the canonical model in which ordinary losses and a rare high-surprisal loss shelf are the only two loss-side information levels.

Proof. The centered high-shelf value is

$$y_1 - h_- = (1-q)(y_1 - y_0) = a,$$

while the centered ordinary-shelf value is

$$y_0 - h_- = -q(y_1 - y_0) < a.$$

Therefore $\{Y - h_- \ge a\}$ is precisely the high shelf and has probability q. Since

$$V_- = q(1-q)(y_1 - y_0)^2, \qquad a^2 = (1-q)^2(y_1 - y_0)^2,$$

we get

$$\frac{V_-}{V_- + a^2} = \frac{q(1-q)(y_1 - y_0)^2}{q(1-q)(y_1 - y_0)^2 + (1-q)^2(y_1 - y_0)^2} = q.$$

Multiplying by p_- gives the unconditional statement. □

Theorem III.11 (Sharp distribution-free high-information bound)**.** *Assume* $0 < V_- < \infty$*. For every* $a > 0$*,*

$$\mathbb{P}(A_a^-) \le p_- \frac{V_-}{V_- + a^2}. \tag{34}$$

The bound uses no monotonicity, tail-index, moment-generating-function, or parametric assumption. In the degenerate case $V_- = 0$*, the right side is interpreted in the limiting sense:* $\iota(X) = h_-$ *almost surely conditional on* $X < 0$*, so* $\mathbb{P}(A_a^-) = 0$ *for every* $a > 0$*.*

Proof. Let $Y = \iota(X)$ under the conditional law $X < 0$. Then $\mathbb{E}Y = h_-$ and $\mathrm{Var}(Y) = V_-$. Cantelli's one-sided Chebyshev inequality [7] gives

$$\mathbb{P}(Y - h_- \ge a \mid X < 0) \le \frac{V_-}{V_- + a^2}, \qquad a > 0.$$

Multiplying by p_- gives (34). □

Proposition III.12 (Optimality of the Cantelli information bound)**.** *Fix* $p_- \in (0,1]$*,* $h_- \in \mathbb{R}$*,* $V_- > 0$*, and* $a > 0$*. Among all laws whose loss-side self-information variable* $Y = \iota(X) \mid X < 0$ *has mean* h_- *and variance* V_-*, no distribution-free upper bound on* $\mathbb{P}(A_a^-)$ *smaller than* (34) *is valid. Equivalently, Cantelli is the best possible bound that uses only* (p_-, h_-, V_-, a)*.*

Proof. It suffices to prove sharpness for the conditional variable Y. Put

$$r := \frac{V_-}{V_- + a^2}, \qquad y_1 := h_- + a, \qquad y_0 := h_- - \frac{V_-}{a}.$$

Let Y take value y_1 with probability r and value y_0 with probability $1 - r$. Then

$$\mathbb{E}Y = h_-, \qquad \mathrm{Var}(Y) = V_-,$$

and

$$\mathbb{P}(Y - h_- \geq a) = r = \frac{V_-}{V_- + a^2}.$$

Thus any smaller bound would fail for a two-point loss-side information law. This two-point score law can be realized by an absolutely continuous density without imposing any geometric regularity: choose two disjoint intervals on the loss side, set the density equal to e^{-y_0} on one and e^{-y_1} on the other, and choose their lengths so that their unconditional masses are $p_-(1-r)$ and $p_- r$. Put the remaining mass $1 - p_-$ on the positive side. Smooth approximations to these two nearly flat information shelves give the same supremum. Hence the obstruction is intrinsic to the chosen summary statistics, not an artifact of allowing discrete score variables. □

The joint magnitude-information profile can now be bounded without pointwise density estimates. For $u \geq 0$ and $a > 0$, define

$$K_-(u, h_- + a) = \mathbb{P}\{X \leq -u,\ \iota(X) - h_- \geq a\}. \tag{35}$$

Then the intersection is no larger than either marginal event.

Corollary III.13 (Distribution-free joint-profile bound)**.** *For every $u \geq 0$ and $a > 0$,*

$$K_-(u, h_- + a) \leq \min\left\{ \bar{F}_-(u), p_- \frac{V_-}{V_- + a^2} \right\}. \tag{36}$$

This bound uses only the empirical magnitude tail and the first two loss-side information moments.

Proof. The event in (35) is contained both in $\{X \leq -u\}$ and in A_a^-. The first containment gives the bound by $\bar{F}_-(u)$; the second gives the bound by Theorem III.11.

□

Theorem III.14 (Joint-profile envelope and summary-level sharpness)**.** *Fix $p_- \in (0, 1]$, $q_u \in [0, p_-]$, $h_- \in \mathbb{R}$, $V_- > 0$, and $a > 0$. Write*

$$r_a := \frac{V_-}{V_- + a^2}.$$

For every absolutely continuous law with

$$\mathbb{P}(X < 0) = p_-, \qquad \mathbb{P}(X \leq -u) = q_u, \qquad \mathbb{E}[\iota(X) \mid X < 0] = h_-, \qquad \mathrm{Var}(\iota(X) \mid X < 0) = V_-,$$

one has

$$K_-(u, h_- + a) \leq \min\{q_u, p_- r_a\}. \tag{37}$$

The envelope is sharp at the level of these summaries in the following precise sense. In the abstract summary model in which only the loss mass p_-, the magnitude-cell mass q_u, and the first two moments of the loss-side score are specified, the largest possible value of the joint cell $\mathbb{P}\{X \leq -u, \iota(X) - h_- \geq a\}$ is $\min\{q_u, p_- r_a\}$. Hence no bound that uses only (p_-, q_u, h_-, V_-, a) can improve on (37). If one also fixes the coordinate geometry of the loss side, such as support lengths, monotonicity, or direct cell counts, then sharper bounds may be possible; those are additional assumptions beyond the summary information used here.

Proof. The upper bound is Theorem III.13 with $q_u = \bar{F}_-(u)$.

For summary-level sharpness, let the conditional loss-side score take the two Cantelli extremal values

$$y_1 = h_- + a, \qquad y_0 = h_- - \frac{V_-}{a},$$

with conditional probabilities r_a and $1 - r_a$. This score law has mean h_-, variance V_-, and unconditional high-score loss mass $p_- r_a$. In an abstract loss-side probability space, choose a magnitude cell M_u with mass q_u and a high-score cell H_a with mass $p_- r_a$, placing the smaller cell inside the larger. Then

$$\mathbb{P}(M_u \cap H_a) = \min\{q_u, p_- r_a\},$$

which equals the upper bound.

Piecewise-constant information-shelf densities realize this construction on the real line whenever the required shelves can be embedded in the chosen loss-side coordinate sets. If the fixed coordinate geometry prevents such an embedding, that obstruction is precisely extra geometric information not present in the summary vector (p_-, q_u, h_-, V_-, a). Thus the distribution-free envelope is sharp for the advertised information set, while any improvement must use more than those summaries. □

For the path certificate, the most useful adverse set is the union

$$A^{\cup}_{u,a} := \{X \le -u\} \cup \{X < 0 : \imath(X) - h_- \ge a\}. \tag{38}$$

It captures large raw losses and high-information loss-side surprises. Its probability is bounded by

$$p^{\cup}_{u,a} := \mathbb{P}(X \in A^{\cup}_{u,a}) \le r_C(u,a) := \left(\bar{F}_-(u) + p_- \frac{V_-}{V_- + a^2}\right) \wedge p_-. \tag{39}$$

The complement of $A^{\cup}_{u,a}$ excludes all one-step losses below $-u$, so the body term is bounded by the ordinary magnitude-truncated Chernoff bound

$$\widehat{B}_T(c,u) := \inf_{\theta \ge 0} e^{-\theta c} \sum_{k=1}^{T} m_u(\theta)^k, \qquad m_u(\theta) = \mathbb{E}\left[e^{-\theta X} \mathbf{1}_{\{X > -u\}}\right]. \tag{40}$$

Theorem III.15 (Practical distribution-free path certificate)**.** *For every $T \ge 1$, floor $c > 0$, magnitude cutoff $u \ge 0$, and information excess threshold $a > 0$,*

$$\mathbb{P}(\underline{S}_T \le -c) \le \widehat{B}_T(c,u) + 1 - \left(1 - r_C(u,a)\right)^T. \tag{41}$$

Consequently,

$$\mathbb{P}(\underline{S}_T \le -c) \le \widehat{B}_T(c,u) + T r_C(u,a). \tag{42}$$

Proof. Apply the set-indexed decomposition to $A = A^{\cup}_{u,a}$:

$$\mathbb{P}(\underline{S}_T \le -c) \le B_T(c; A^{\cup}_{u,a}) + H_T(A^{\cup}_{u,a}).$$

Since $A^{\cup}_{u,a} \supseteq \{X \le -u\}$, avoiding $A^{\cup}_{u,a}$ implies avoiding all losses below $-u$. Hence

$$B_T(c; A^{\cup}_{u,a}) \le B_T(c; \{X \le -u\}) \le \widehat{B}_T(c,u).$$

Also $\mathbb{P}(X \in A^{\cup}_{u,a}) \le r_C(u,a)$, so

$$H_T(A^{\cup}_{u,a}) = 1 - \left(1 - \mathbb{P}(X \in A^{\cup}_{u,a})\right)^T \le 1 - (1 - r_C(u,a))^T.$$

This proves (41). The linearized bound follows from $1 - (1-r)^T \le Tr$.

□

Theorem III.16 (Hit-budget envelope and summary-level sharpness)**.** *Fix $T \ge 1$, $p_- \in (0,1]$, $q_u \in [0, p_-]$, $h_- \in \mathbb{R}$, $V_- > 0$, and $a > 0$. Let*

$$r_a := \frac{V_-}{V_- + a^2}, \qquad r_* := \min\{p_-, q_u + p_- r_a\}.$$

For the union adverse set

$$A^{\cup}_{u,a} = \{X \le -u\} \cup \{X < 0 : \imath(X) - h_- \ge a\},$$

every law with $\mathbb{P}(X < 0) = p_-$, $\mathbb{P}(X \le -u) = q_u$, and loss-side score mean and variance (h_-, V_-) satisfies

$$H_T(A^{\cup}_{u,a}) \le 1 - (1 - r_*)^T. \tag{43}$$

At the same summary level this hit budget is sharp: in the abstract summary model using only $(p_-, q_u, h_-, V_-, a, T)$, the right side of (43) is attainable. Thus no smaller distribution-free hit budget can be justified from those summaries alone. Sharper coordinate-level bounds require additional information about the overlap between the magnitude cell and the high-score cell.

Proof. By Cantelli,

$$\mathbb{P}\{X < 0 : \imath(X) - h_- \geq a\} \leq p_- r_a.$$

Therefore the one-period union probability is bounded by

$$\mathbb{P}(A^{\cup}_{u,a}) \leq \min\{p_-, q_u + p_- r_a\} = r_*.$$

Independence over time gives

$$H_T(A^{\cup}_{u,a}) = 1 - \left(1 - \mathbb{P}(A^{\cup}_{u,a})\right)^T \leq 1 - (1 - r_*)^T.$$

For summary-level sharpness, use the two-point Cantelli extremal score law from Theorem III.12, so the high-score loss cell has unconditional mass $p_- r_a$. Choose the overlap between the magnitude cell $M_u = \{X \leq -u\}$ and the high-score cell H_a to be as small as the loss-side mass constraint allows:

$$\mathbb{P}(M_u \cap H_a) = \left(q_u + p_- r_a - p_-\right)_+.$$

Then

$$\mathbb{P}(M_u \cup H_a) = q_u + p_- r_a - \left(q_u + p_- r_a - p_-\right)_+ = r_*.$$

The corresponding T-period hit probability is exactly $1-(1-r_*)^T$. As in Theorem III.14, piecewise-constant information shelves realize the construction whenever the relevant coordinate cells can be embedded; absent such geometric information, the summary-level hit budget cannot be improved. □

Theorem III.17 (Profile-to-floor-breach transfer)**.** *Fix a floor $c > 0$ and a single i.i.d. increment law. For each horizon T, choose $b_T \geq 0$ and a score threshold t_T, and define*

$$A_T := \{X \leq -(c + b_T),\ \imath(X) \geq t_T\}, \qquad p_T := \mathbb{P}(X \in A_T) = K_-(c + b_T, t_T).$$

Assume that, as $T \to \infty$,

$$T p_T \to \lambda \in [0, \infty), \qquad B_T(c; A_T) \to 0, \qquad \mathbb{P}\left(\max_{0 \leq k < T} S_k > b_T\right) \to 0. \tag{44}$$

Then

$$\mathbb{P}(\underline{S}_T \leq -c) \longrightarrow 1 - e^{-\lambda}. \tag{45}$$

Thus profile mass at the $1/T$ scale becomes order-one finite-horizon floor-breach risk whenever the selected profile cell is breach-effective and the body mechanism is negligible.

Proof. The set-indexed upper bound gives

$$\mathbb{P}(\underline{S}_T \leq -c) \leq B_T(c; A_T) + 1 - (1 - p_T)^T.$$

By assumption, $B_T(c; A_T) \to 0$ and $T p_T \to \lambda$, so the upper limit is at most $1 - e^{-\lambda}$.

For the lower bound, let $D_T(A_T) = \{\exists i \leq T : X_i \in A_T\}$. On the event

$$D_T(A_T) \cap \left\{\max_{0 \leq k < T} S_k \leq b_T\right\},$$

let $\tau \leq T$ be the first time with $X_\tau \in A_T$. Then

$$S_\tau = S_{\tau - 1} + X_\tau \leq b_T - (c + b_T) = -c,$$

so $\underline{S}_T \leq -c$. Hence

$$\mathbb{P}(\underline{S}_T \leq -c) \geq \mathbb{P}(D_T(A_T)) - \mathbb{P}\left(\max_{0 \leq k < T} S_k > b_T\right).$$

The second term tends to zero by assumption, while $\mathbb{P}(D_T(A_T)) = 1 - (1 - p_T)^T \to 1 - e^{-\lambda}$. The lower and upper bounds match. □

The transfer theorem above is a qualitative limit: profile mass at the $1/T$ scale becomes order-one breach risk $1-e^{-\lambda}$. Its proof in fact yields an exact two-sided sandwich at every finite horizon, and the limiting form is one explicit Poisson-approximation step away. We record the non-asymptotic version, from which Theorem III.17 follows as a corollary. Three error mechanisms appear, each separately controllable: the body term, the path-overshoot term, and the discrete Poisson-approximation error of the adverse-hit count.

Fix the increment law, a floor $c>0$, a horizon $T\in\mathbb{N}$, a buffer $b_T\geq 0$, and a score threshold t_T. As in Theorem III.17 set

$$A_T := \{X\leq -(c+b_T),\ \iota(X)\geq t_T\},\qquad p_T := \mathbb{P}(X\in A_T) = K_{-}(c+b_T,\ t_T),\qquad \lambda_T := T p_T.$$

Write $H_T := H_T(A_T) = 1-(1-p_T)^T$ for the exact adverse-hit probability, $B_T := B_T(c;A_T)$ for the body breach probability on paths avoiding A_T, and

$$M_T := \mathbb{P}\Big(\max_{0\leq k<T} S_k > b_T\Big)$$

for the path-overshoot probability.

Theorem III.18 (Quantitative profile-to-floor-breach transfer)**.** *With the notation above, the following hold for every finite T.*

1. (Exact sandwich.)
$$H_T - M_T \ \leq\ \mathbb{P}(\underline{S}_T\leq -c)\ \leq\ H_T + B_T. \tag{46}$$
2. (Poisson rate.) *The adverse-hit count $N_T=\sum_{i=1}^{T}\mathbf{1}\{X_i\in A_T\}\sim \mathrm{Binomial}(T,p_T)$ satisfies*
$$\Big|H_T-(1-e^{-\lambda_T})\Big|\ \leq\ d_{\mathrm{TV}}\big(N_T,\ \mathrm{Poisson}(\lambda_T)\big)\ \leq\ \min(\lambda_T,1)\,p_T\ \leq\ \frac{\lambda_T^2}{T}. \tag{47}$$
3. (Quantitative transfer.) *Consequently,*
$$\Big|\mathbb{P}(\underline{S}_T\leq -c)-(1-e^{-\lambda_T})\Big|\ \leq\ B_T+M_T+\min(\lambda_T,1)\,p_T, \tag{48}$$
and, comparing against a fixed target $\lambda\geq 0$,
$$\Big|\mathbb{P}(\underline{S}_T\leq -c)-(1-e^{-\lambda})\Big|\ \leq\ B_T+M_T+\min(\lambda_T,1)\,p_T+|\lambda_T-\lambda|. \tag{49}$$
4. (Explicit error control.) *If A_T is Chernoff-admissible at some $\theta>0$, then the body term obeys the generic set-indexed Chernoff body bound $B_T\leq \widehat{B}_T(c;A_T)$ from Theorem III.3 and (6). If, in addition, $m(\theta):=\mathbb{E}[e^{\theta X}]<\infty$ for some $\theta>0$, then for every such θ,*
$$M_T\ \leq\ e^{-\theta b_T}\,\big[\max\{1,m(\theta)\}\big]^{T-1}, \tag{50}$$
and in particular, when $\mathbb{E}X\geq 0$, $M_T\leq e^{-\theta b_T}\,m(\theta)^{T-1}$.

Proof. (1) Exact sandwich. The upper bound is the set-indexed decomposition Theorem III.3 applied to A_T: $\mathbb{P}(\underline{S}_T\leq -c)\leq B_T(c;A_T)+H_T(A_T)$. For the lower bound, let $\tau := \inf\{i\geq 1 : X_i\in A_T\}$ be the first adverse hit, and work on the event $\{\tau\leq T\}\cap\{\max_{0\leq k<T} S_k\leq b_T\}$. On this event $S_{\tau-1}\leq \max_{0\leq k\leq T-1} S_k\leq b_T$ because $\tau-1\leq T-1$, and $X_\tau\leq -(c+b_T)$ because $X_\tau\in A_T$. Hence

$$S_\tau = S_{\tau-1}+X_\tau\leq b_T-(c+b_T) = -c,$$

so $\underline{S}_T\leq -c$. Therefore

$$\mathbb{P}(\underline{S}_T\leq -c)\ \geq\ \mathbb{P}\Big(\{\tau\leq T\}\cap\{\max_{0\leq k<T} S_k\leq b_T\}\Big)\ \geq\ \mathbb{P}(\tau\leq T)-M_T\ =\ H_T-M_T,$$

using $\mathbb{P}(\tau\leq T)=\mathbb{P}(N_T\geq 1)=H_T$ and a union bound on the complement of $\{\max_{k<T} S_k\leq b_T\}$.

(2) Poisson rate. By independence, N_T is $\mathrm{Binomial}(T,p_T)$, the sum of T i.i.d. Bernoulli(p_T) indicators. The event $\{N_T\geq 1\}$ is a subset of $\mathbb{Z}_{\geq 0}$; for any such event the gap between the N_T-probability and the $\mathrm{Poisson}(\lambda_T)$-probability is at most the total variation distance, so

$$\Big|\mathbb{P}(N_T\geq 1)-\mathbb{P}(\mathrm{Poisson}(\lambda_T)\geq 1)\Big|\leq d_{\mathrm{TV}}\big(N_T,\mathrm{Poisson}(\lambda_T)\big).$$

The left side is $|H_T-(1-e^{-\lambda_T})|$. The Le Cam bound [8] gives $d_{\mathrm{TV}}\leq\sum_{i=1}^{T}p_T^2=Tp_T^2=\lambda_T^2/T$, and the Barbour–Hall refinement [9] sharpens the constant to $d_{\mathrm{TV}}\leq\min(1,\lambda_T^{-1})\sum_i p_T^2=\min(\lambda_T,1)\,p_T$. This is (47).

(3) Quantitative transfer. Combining (46) with (47): from the sandwich, $-M_T \le \mathbb{P}(\underline{S}_T \le -c) - H_T \le B_T$; adding and subtracting $(1-e^{-\lambda_T})$ and using (47) to replace H_T up to the stated error gives

$$-\big(M_T + \min(\lambda_T, 1)p_T\big) \le \mathbb{P}(\underline{S}_T \le -c) - (1-e^{-\lambda_T}) \le B_T + \min(\lambda_T, 1)p_T,$$

which is (48). Inequality (49) follows from the triangle inequality and the fact that $x \mapsto e^{-x}$ is 1-Lipschitz on $[0,\infty)$, so $|e^{-\lambda_T} - e^{-\lambda}| \le |\lambda_T - \lambda|$.

(4) Explicit error control. The bound $B_T \le \widehat{B}_T(c; A_T)$ is the generic set-indexed Chernoff body bound from Theorem III.3 and (6), valid whenever A_T is Chernoff-admissible. For the overshoot term, fix $\theta > 0$ with $m(\theta) < \infty$. If $m(\theta) \ge 1$, then $\{e^{\theta S_k}\}_{k\ge 0}$ is a nonnegative submartingale, because $\mathbb{E}[e^{\theta S_{k+1}} \mid \mathcal{F}_k] = e^{\theta S_k} m(\theta) \ge e^{\theta S_k}$; Doob's maximal inequality applied to $e^{\theta S_0}, \dots, e^{\theta S_{T-1}}$ gives

$$M_T = \mathbb{P}\Big(\max_{0\le k\le T-1} S_k > b_T\Big) \le e^{-\theta b_T}\, \mathbb{E}[e^{\theta S_{T-1}}] = e^{-\theta b_T}\, m(\theta)^{T-1}.$$

If instead $m(\theta) < 1$, then $\{e^{\theta S_k} m(\theta)^{-k}\}$ is a nonnegative martingale of unit mean, and since $m(\theta)^{-k} \ge 1$ we have $e^{\theta S_k} \le e^{\theta S_k} m(\theta)^{-k}$; Ville's maximal inequality for the martingale then yields $\mathbb{P}(\max_{k\le T-1} S_k > b_T) \le e^{-\theta b_T}$. Both cases are summarized by the single bound (50) with $\max\{1, m(\theta)\}$. When $\mathbb{E}X \ge 0$, Jensen gives $m(\theta) \ge e^{\theta \mathbb{E}X} \ge 1$, so the first case always applies and the exponent base is $m(\theta)$. □

Corollary III.19 (Recovery of the qualitative transfer theorem)**.** *Suppose, as $T \to \infty$, that $\lambda_T \to \lambda \in [0,\infty)$, $B_T \to 0$, and $M_T \to 0$. Then $\mathbb{P}(\underline{S}_T \le -c) \to 1 - e^{-\lambda}$.*

Proof. In (49) the term $\min(\lambda_T, 1)p_T \le \lambda_T^2/T \to 0$ since λ_T is bounded, and $|\lambda_T - \lambda| \to 0$ by hypothesis, while $B_T \to 0$ and $M_T \to 0$. Hence the right side tends to 0, which is the conclusion of Theorem III.17. □

Remark III.20*.* Three points distinguish Theorem III.18 from the limit it contains. First, (46) is exact and assumption-free: it holds for any horizon and any increment law, with the limiting Poisson form isolated as the single discrete step (47). Second, every error term is a quantity the certificate already estimates: B_T is the Chernoff body bound, M_T is an overshoot probability with the explicit MGF bound (50), and the Poisson term is λ_T^2/T, governed by the same profile mass $p_T = K_-(c + b_T, t_T)$ that drives the first-order limit. The leading correction to $1 - e^{-\lambda}$ is therefore $O(1/T)$ in the hit count, with the body and overshoot terms controlling the finite-horizon validity of the single-big-jump picture. Third, the buffer b_T is now a visible tradeoff parameter: increasing b_T deepens the adverse shelf (lowering p_T, hence λ_T) but shrinks the overshoot term M_T through (50), exactly the body-versus-hit balance that the set-indexed certificate is built to expose.

Proposition III.21 (Score-filtered body refinement)**.** *If log-scores are available, the body bound in Theorem III.15 can be sharpened without adding a parametric assumption. Define*

$$m^{\cup}_{u,a}(\theta) := \mathbb{E}\Big[e^{-\theta X}\, \mathbf{1}_{\{X \notin A^{\cup}_{u,a}\}}\Big] = \mathbb{E}\Big[e^{-\theta X}\, \mathbf{1}_{\{X > -u\}} \mathbf{1}_{\{X \ge 0 \text{ or } \iota(X) - h_- < a\}}\Big], \tag{51}$$

and

$$\widehat{B}^{\cup}_T(c, u, a) := \inf_{\theta \ge 0} e^{-\theta c} \sum_{k=1}^{T} \big(m^{\cup}_{u,a}(\theta)\big)^k. \tag{52}$$

Then

$$\mathbb{P}(\underline{S}_T \le -c) \le \widehat{B}^{\cup}_T(c, u, a) + 1 - \big(1 - r_C(u, a)\big)^T. \tag{53}$$

Moreover $\widehat{B}^{\cup}_T(c, u, a) \le \widehat{B}_T(c, u)$, so (53) improves the conservative certificate whenever high-information loss states materially contribute to the body bound.

Proof. The proof is the same Chernoff argument as in Theorem III.15, now applied directly to the adverse set $A^{\cup}_{u,a}$. Since $m^{\cup}_{u,a}(\theta) \le m_u(\theta)$ for every $\theta \ge 0$, the displayed monotonicity of the body bounds follows. □

Corollary III.22 (Distribution-free hit budget)**.** *Let $\delta_T(\alpha) := 1 - (1-\alpha)^{1/T}$. Suppose a mandate splits the breach budget as $\alpha = \alpha_{\text{body}} + \alpha_{\text{hit}}$. The conditions*

$$\widehat{B}_T(c, u) \le \alpha_{\text{body}} \tag{54}$$

and

$$r_C(u, a) \le \delta_T(\alpha_{\text{hit}}) \tag{55}$$

certify $\mathbb{P}(\underline{S}_T \le -c) \le \alpha$. If $p_- \le \delta_T(\alpha_{\rm hit})$, the hit budget is automatic. Otherwise, if $\bar{F}_-(u) < \delta_T(\alpha_{\rm hit})$, it is certified by

$$a^2 \ge V_- \left(\frac{p_-}{\delta_T(\alpha_{\rm hit}) - \bar{F}_-(u)} - 1 \right)_+ . \tag{56}$$

The rare-hit approximation replaces $\delta_T(\alpha_{\rm hit})$ by $\alpha_{\rm hit}/T$.

Proof. By Theorem III.15, the body and hit conditions imply

$$\mathbb{P}(\underline{S}_T \le -c) \le \alpha_{\rm body} + 1 - (1 - r_C(u,a))^T \le \alpha_{\rm body} + \alpha_{\rm hit}.$$

The algebraic condition (56) is just

$$\bar{F}_-(u) + p_- \frac{V_-}{V_- + a^2} \le \delta_T(\alpha_{\rm hit}),$$

solved for a^2. □

Remark III.23 (What can improve the bound)*.* The optimality result in Theorem III.12 is conditional on the chosen statistically estimable summaries. No sharper distribution-free bound is available from (p_-, h_-, V_-) alone. A sharper certificate must use additional information: higher moments of the loss-side self-information score, direct empirical counts of the cell $\{X \le -u, \iota(X) - h_- \ge a\}$, or a geometric assumption on the loss-side density. The next subsection records a simple magnitude-information refinement. It is optional; the practical certificate above does not require it.

III-F Information-moment limitations: a converse and a score-cumulant separation result

Theorems III.12 and III.14 show that the Cantelli hit budget is the sharp distribution-free bound when the available information is only (p_-, h_-, V_-), together with the magnitude-cell mass in the joint-profile case. We now record what loss-side information moments can and cannot do, and we separate two questions that are easily conflated.

Identification (one fixed law). Is the finite-horizon breach probability $P(\underline{S}_T \le -c)$ a functional of the loss-side score law $\mathcal{L}(Y)$, $Y = \iota(X) \mid \{X < 0\}$? Theorem III.24 answers no: two increment laws can share the *entire* conditional loss-side score law—hence h_-, V_-, and every higher score cumulant—yet have different breach probabilities. The mechanism is a location invariance: on a uniform shelf of mass p and width ε the score is $y = \log(\varepsilon/p)$, unchanged by translating the shelf along the magnitude axis. The score law therefore quotients out exactly the magnitude-location information on which breach risk depends, so breach risk varies along the fibers of that quotient and is pinned down by no list of score moments.

Discrimination (two laws). If two laws have *different* score laws, can the difference be detected from data? Theorem III.26 answers yes: a fixed gap in any estimable score cumulant—we use the third—is detected at the parametric rate, with power tending to one.

These statements do not conflict, because Theorem III.26 presupposes precisely the situation Theorem III.24 excludes. The converse pair has *identical* score laws, so Theorem III.26 is silent on it; and detecting a score difference does not pin down breach risk, since the score-to-breach map is not a function. The conclusion is the same from both directions: the primitive object must be the two-dimensional profile $K_-(u,t)$, which retains the magnitude coordinate, rather than any list of one-dimensional score moments.

The statement is intentionally formulated with an embeddability condition. A piecewise-constant score law fixes not only shelf probabilities but also shelf lengths, because a shelf of mass p and score y has length pe^y. If one also imposes a fixed floor and a fixed loss half-line, coordinate geometry can impose additional restrictions. The converse below isolates the structural point in the regime where the same score shelves can be embedded either as shallow losses or as breach-capable losses.

Theorem III.24 (Converse: score moments do not locate adverse mass)**.** *Fix a horizon $T \ge 2$ and a floor $c > 0$. Let a nondegenerate two-shelf loss-side information law be specified by loss masses $p_o, p_d > 0$, widths $\varepsilon_o, \varepsilon_d > 0$, and $p_o + p_d = p_- \le 1$. Suppose these two loss shelves can be embedded inside the shallow interval $(-c/T, 0)$, i.e.*

$$\varepsilon_o + \varepsilon_d < c/T,$$

with a positive gap between shelves. Then there exist two absolutely continuous increment laws P and Q with the same loss mass and the same full conditional loss-side information law, hence the same h_-, V_-, and all higher loss-side information cumulants, but with different finite-horizon floor breach probabilities:

$$P(\underline{S}_T \le -c) \ne Q(\underline{S}_T \le -c).$$

Consequently no functional of the loss-side information-score law alone can determine finite-horizon floor-breach risk. Magnitude information, such as $\bar{F}_-(u)$ or the full profile $K_-(u,t)$, is indispensable.

Proof. On a uniform shelf of center x, width ε, and mass p, the density is p/ε and the self-information is the constant

$$y = -\log(p/\varepsilon) = \log(\varepsilon/p),$$

which does not depend on the shelf location x. Thus translating a shelf changes its economic magnitude but leaves the loss-side score value unchanged.

Construct Q by placing both loss shelves inside $(-c/T, 0)$, using the specified masses and widths. Put any remaining mass $1 - p_-$, if present, on a fixed positive shelf. Then every one-period loss under Q has magnitude less than c/T, so even T consecutive loss increments cannot cross the floor; therefore

$$Q(\underline{S}_T \le -c) = 0.$$

Construct P with the same shelf masses and widths and the same positive shelf, but translate the shelf of mass p_d to an interval $J_d \subset (-\infty, -c)$. If the translated shelf is hit at the first period, then the floor is crossed immediately: on $\{X_1 \in J_d\}$, $S_1 = X_1 < -c$, and hence $\underline{S}_T \le -c$. Therefore

$$P(\underline{S}_T \le -c) \ge P(X_1 \in J_d) = p_d > 0.$$

This weaker lower bound is all that is needed to separate the two laws. We do not claim that an arbitrary later hit of J_d necessarily breaches the floor, because gains accumulated before the hit may buffer the loss. □

Remark III.25 (When the stronger hit lower bound is valid). The stronger lower bound $1 - (1 - p_d)^T$ is valid only under an additional buffer condition. For example, if all nontranslated increments are bounded above by $G < \infty$ and the translated shelf is placed inside

$$(-\infty, -c - (T-1)G],$$

then any hit of that shelf before time T breaches the floor regardless of the gains accumulated earlier. Without such a condition, the proof above uses the first-period hit event and obtains the sufficient lower bound $p_d > 0$.

As flagged above, the next theorem answers the discrimination question, not the identification one. Its hypothesis is that the two score laws genuinely differ; it therefore says nothing about the converse pair of Theorem III.24, whose score laws coincide. It is an achievability counterpart—useful for score-distinguished model classes—and not a converse to Theorem III.24.

Theorem III.26 (Third-cumulant separation for score-distinguished families). *Let $\mathcal{F}$ be a family of conditional loss-side information laws with common mean h_- and variance V_-, but with possibly different third central information moments*

$$\kappa_3^- := \mathbb{E}[(Y - h_-)^3], \qquad Y = \iota(X) \mid \{X < 0\}.$$

Assume every member of the family has a finite sixth central information moment. For two members $P, Q \in \mathcal{F}$ with $\kappa_3^-(P) \ne \kappa_3^-(Q)$, let $\hat{\kappa}_{3,n}^-$ be the sample third central moment computed from n_- loss-side score observations. Then, under each fixed law,

$$\sqrt{n_-}\left(\hat{\kappa}_{3,n}^- - \kappa_3^-\right) \Rightarrow N(0, \sigma_3^2),$$

where, writing $\mu_j = \mathbb{E}[(Y - h_-)^j]$,

$$\sigma_3^2 = \mu_6 - \mu_3^2 - 6\mu_4\mu_2 + 9\mu_2^3.$$

Consequently the standardized two-law separation statistic for independent samples satisfies

$$\frac{\mathbb{E}[\hat{\kappa}_{3,n}^-(P) - \hat{\kappa}_{3,n}^-(Q)]}{\mathrm{sd}(\hat{\kappa}_{3,n}^-(P) - \hat{\kappa}_{3,n}^-(Q))} = \Theta(\sqrt{n_-}),$$

so any fixed third-cumulant gap is detected with power tending to one.

Proof. The sample third central moment is a smooth polynomial function of the first three raw sample moments. Under the finite-sixth-moment assumption, the raw moment vector satisfies a multivariate central limit theorem. Applying the delta method to the centered third-moment functional gives the displayed normal limit and the classical variance formula; see, for example, [10, Chapter 3]. For two independent samples, the difference of the two estimators is asymptotically normal with fixed nonzero mean $\kappa_3^-(P) - \kappa_3^-(Q)$ and standard deviation of order $n_-^{-1/2}$. The standardized separation is therefore of order $\sqrt{n_-}$, and a fixed normal-quantile test has power tending to one. □

Remark III.27 (Relation to the controlled experiment). The controlled experiment in Section IV-D uses a pair already separated by the second score cumulant V_-, so it lives in the discrimination regime of Theorem III.26, not the identical-score regime of Theorem III.24. Neither result implies that the location deficiency of Theorem III.24 closes at any finite cumulant order: that deficiency is intrinsic to the score law and is removed only by adding magnitude information through $K_-(u,t)$.

III-G Optional geometric refinement: magnitude-information growth

The baseline certificate intentionally avoids pointwise density levels and tail indices. A practitioner may nevertheless be willing to impose a weak geometric condition on the loss side. The condition must be stated as a tail-uniform information floor, not as a pointwise statement at the boundary. Let

$$I_-(u) := \operatorname*{ess\,inf}_{x \le -u} \imath(x).$$

For a scale $s_\rho > 0$, suppose that sufficiently large losses force at least $\rho(u/s_\rho)$ excess surprisal:

$$I_-(u) - h_- \ge \rho\left(\frac{u}{s_\rho}\right), \qquad u \ge u_0, \tag{57}$$

where ρ is nondecreasing. This is not a parametric return model; it is a magnitude-to-information envelope. It says that the whole loss tail beyond $-u$, up to null sets, cannot sit in an unusually high-density pocket. If the loss-side density is monotone as one moves left, then $I_-(u) = \imath(-u)$ and the older boundary notation is recovered. Without such monotonicity, the essential infimum in (57) is the required population object.

Under (57), every $X \le -u$ satisfies $\imath(X) - h_- \ge \rho(u/s_\rho)$ almost surely. Hence any joint event $\{X \le -u, \imath(X) - h_- \ge a\}$ implies the single score event

$$\imath(X) - h_- \ge a_\rho(u,a), \qquad a_\rho(u,a) := \max\left\{a, \rho\left(\frac{u}{s_\rho}\right)\right\}.$$

Therefore the Cantelli branch improves to

$$K_-(u, h_- + a) \le p_- \frac{V_-}{V_- + a_\rho(u,a)^2}. \tag{58}$$

If a practitioner also specifies a one-dimensional score-tail envelope

$$\mathbb{P}\{\imath(X) - h_- \ge r \mid X < 0\} \le \Phi(r), \tag{59}$$

then (58) is replaced by

$$K_-(u, h_- + a) \le p_- \Phi(a_\rho(u,a)). \tag{60}$$

Loss-side information regularity does not improve the pure high-information bound by itself; Cantelli remains optimal from (p_-, h_-, V_-). Its role is to tie economically large losses to the score coordinate when such geometry is credible.

III-H Pareto-score envelopes: a nonparametric score-tail stress parameter

The sharp Cantelli bound is the correct baseline when only (p_-, h_-, V_-) are used. A practitioner may nevertheless be willing to add a weak, one-dimensional regularity assumption about the downturn score tail. This does not require a parametric return model. It specifies only how quickly the conditional excess surprisal

$$\bar{G}_-(a) := \mathbb{P}\{\imath(X) - h_- \ge a \mid X < 0\}, \qquad a \ge 0, \tag{61}$$

decays. We use a separate symbol α_{P} for the Pareto-score benchmark index, to avoid confusion with risk-budget levels such as α_{hit} or VaR levels.

Definition III.28 (Pareto-score envelope)**.** For $\alpha_{\mathrm{P}} > 0$, define

$$\lambda_{\mathrm{P}} := \lambda(\alpha_{\mathrm{P}}) := \frac{\alpha_{\mathrm{P}}}{\alpha_{\mathrm{P}} + 1}. \tag{62}$$

We say that the loss-side score tail is no heavier than the Pareto- α_{P} score benchmark above level a_0 if there is a constant C_{P} such that

$$\bar{G}_{-}(a) \le C_{\mathrm{P}} e^{-\lambda_{\mathrm{P}} a}, \qquad a \ge a_0. \tag{63}$$

The canonical shifted-Pareto normalization is

$$C_{\mathrm{P}} = e^{-1}, \qquad a_0 = 0, \qquad \bar{G}_{-}(a) \le e^{-1-\lambda_{\mathrm{P}} a}. \tag{64}$$

The envelope (63) is not an assumption that returns themselves are Pareto. It is a stress assumption on the one-dimensional loss-side log-density score. Smaller α_{P} means a heavier permitted score tail and therefore a weaker, more conservative regularity assumption. Larger α_{P} means faster score decay and a tighter risk certificate.

Combining this optional envelope with the always-valid Cantelli inequality gives an effective score-tail bound

$$\Phi_{\alpha_{\mathrm{P}}}(a) := \min\left\{ \frac{V_{-}}{V_{-} + a^2}, C_{\mathrm{P}} e^{-\lambda_{\mathrm{P}} a} \right\}, \qquad a \ge a_0. \tag{65}$$

Thus

$$\mathbb{P}\{X < 0,\ \iota(X) - h_{-} \ge a\} \le p_{-} \Phi_{\alpha_{\mathrm{P}}}(a). \tag{66}$$

The distribution-free case is recovered by dropping the second branch.

With the Pareto-score envelope, the joint magnitude-information profile satisfies

$$K_{-}(u, h_{-} + a) \le \min\left\{ \bar{F}_{-}(u), p_{-} \Phi_{\alpha_{\mathrm{P}}}(a) \right\}. \tag{67}$$

For the practical union adverse set $A^{\cup}_{u,a}$, define

$$r_{\alpha_{\mathrm{P}}}(u, a) := \left[\bar{F}_{-}(u) + p_{-} \Phi_{\alpha_{\mathrm{P}}}(a) \right] \wedge p_{-}. \tag{68}$$

Then the path-concentration certificate becomes

$$\mathbb{P}(\underline{S}_T \le -c) \le \widehat{B}^{\cup}_T(c, u, a) + 1 - \left(1 - r_{\alpha_{\mathrm{P}}}(u, a)\right)^T, \tag{69}$$

with the rare-hit version

$$\mathbb{P}(\underline{S}_T \le -c) \le \widehat{B}^{\cup}_T(c, u, a) + T r_{\alpha_{\mathrm{P}}}(u, a). \tag{70}$$

Compared with the fully distribution-free certificate, the only change is the replacement

$$\frac{V_{-}}{V_{-} + a^2} \quad \rightsquigarrow \quad \Phi_{\alpha_{\mathrm{P}}}(a).$$

The body-bound term is unchanged; only the score-hit budget is tightened.

α_{P}	λ_{P}	$a = 2$	$a = 4$	$a = 6$
1	1/2	0.1353	0.0498	0.0183
2	2/3	0.0969	0.0256	0.0067
4	4/5	0.0743	0.0150	0.0030
∞	1	0.0498	0.0067	0.0009

Table 1. Canonical Pareto-score benchmark probabilities $e^{-1-\lambda_{\mathrm{P}} a}$ for raw excess-surprisal thresholds a measured in nats. The index α_{P} is a score-tail stress parameter, not a return-tail parameter. Larger α_{P} imposes faster downturn-score decay and gives a tighter hit-risk certificate.

The risk-budget algebra shows the practical improvement. Let $\delta_T(\alpha_{\mathrm{hit}}) = 1 - (1 - \alpha_{\mathrm{hit}})^{1/T}$. The hit budget is certified if

$$\bar{F}_{-}(u) + p_{-} \Phi_{\alpha_{\mathrm{P}}}(a) \le \delta_T(\alpha_{\mathrm{hit}}). \tag{71}$$

When the canonical Pareto-score branch is active, this requires only

$$a \ge \frac{1}{\lambda_{\mathrm{P}}}\left[\log\left(\frac{p_-}{\delta_T(\alpha_{\mathrm{hit}}) - \bar{F}_-(u)}\right) - 1\right], \tag{72}$$

provided $\bar{F}_-(u) < \delta_T(\alpha_{\mathrm{hit}})$. The Cantelli-only condition instead requires

$$a^2 \ge V_-\left(\frac{p_-}{\delta_T(\alpha_{\mathrm{hit}}) - \bar{F}_-(u)} - 1\right). \tag{73}$$

Thus the distribution-free certificate needs a square-root inverse-budget information threshold, while the Pareto-score envelope needs only a logarithmic threshold. This is the value of adding a score-tail regularity assumption: it can materially tighten the hit budget without choosing a parametric model for returns.

A practitioner who wants a two-sided stress band for downturn-score risk can use two Pareto-score indices $0 < \alpha_{\mathrm{L}} \le \alpha_{\mathrm{U}}$. Since $\lambda(\alpha) = \alpha/(\alpha+1)$ is increasing, a canonical band is

$$e^{-1-\lambda(\alpha_{\mathrm{U}})a} \le \bar{G}_-(a) \le e^{-1-\lambda(\alpha_{\mathrm{L}})a}. \tag{74}$$

The upper bound is what enters the conservative path certificate. The lower bound is useful for model validation and stress interpretation, but by itself it is not a lower bound on floor-breach probability unless the high-score downturns are also large enough in magnitude.

III-I Why a Pareto-score envelope is not a regular-variation assumption

The terminology "Pareto-score" is meant literally in score space and not in return space. Regular variation of the loss tail is one sufficient mechanism for a Pareto-score tail, but it is not required.

Let $L = -X > 0$ denote loss magnitude under the conditional loss-side law. If L has an exact Pareto density with index α_{P},

$$f_L(\ell) = \alpha_{\mathrm{P}} x_m^{\alpha_{\mathrm{P}}} \ell^{-(\alpha_{\mathrm{P}}+1)}, \qquad \ell \ge x_m, \tag{75}$$

then

$$Y = -\log f_L(L) = C + (\alpha_{\mathrm{P}} + 1)\log L.$$

Since $\log(L/x_m)$ is exponential with rate α_{P}, the score Y is shifted exponential with rate

$$\lambda_{\mathrm{P}} = \frac{\alpha_{\mathrm{P}}}{\alpha_{\mathrm{P}} + 1},$$

and therefore

$$\mathbb{P}(Y - \mathbb{E}Y \ge a) = e^{-1-\lambda_{\mathrm{P}}a}. \tag{76}$$

More generally, if $f_L(\ell) \sim C\ell^{-(\alpha_{\mathrm{P}}+1)}L_0(\ell)$ with a slowly varying factor L_0, the excess-score tail has logarithmic rate λ_{P}, possibly with slowly varying corrections. This proves only the implication

$$\text{regularly varying loss tail} \quad \Longrightarrow \quad \text{Pareto-score tail with rate } \lambda_{\mathrm{P}}.$$

The converse is false and is not needed.

Bounded-support laws can satisfy Pareto-score envelopes as well. Suppose a conditional loss density on $[0, M]$ vanishes polynomially near the worst-loss endpoint:

$$f_L(M - r) \asymp Cr^q, \qquad r \downarrow 0, \qquad q > 0. \tag{77}$$

Then near the endpoint

$$-\log f_L(M - r) = C_0 - q\log r,$$

while $\mathbb{P}(M - L \le r) \asymp r^{q+1}$. Transforming r into score gives an exponential score tail with rate

$$\lambda_{\mathrm{edge}} = \frac{q+1}{q} > 1. \tag{78}$$

This is faster than every Pareto-score benchmark, whose rate $\alpha_{\mathrm{P}}/(\alpha_{\mathrm{P}}+1)$ is below one. If the density is positive at the endpoint $(q = 0)$, the worst-loss score is bounded rather than exponentially tailed; the magnitude term $\bar{F}_-(u)$, not the score term, then carries the large bounded-loss risk.

This is exactly why the practical adverse set is a union:

$$A^{\cup}_{u,a} = \{X \le -u\} \cup \{X < 0 : \imath(X) - h_- \ge a\}.$$

The magnitude component catches large bounded losses even when they are not low-density score surprises. The score component catches low-density adverse states even when a simple magnitude threshold is not the whole story. A Pareto-score envelope controls the second component. It does not replace the first.

III-J Worked comparison: how conservative is the sharp distribution-free bound?

The optimality result in Theorem III.12 is worst-case. It is therefore useful to ask how far the sharp distribution-free envelope is from classical smooth laws that are common in financial modeling. This comparison is not used to prove the theorem; it calibrates the cost of remaining assumption-light.

For symmetric examples, $p_- = 1/2$, $h_- = h$, and $V_- = \mathrm{VE}(X)$. Choosing an excess-information threshold $a = k\sqrt{V_-}$, the universal bound becomes

$$\mathbb{P}\{X < 0, \imath(X) - h \ge k\sqrt{\mathrm{VE}(X)}\} \le \frac{1}{2(1+k^2)}. \tag{79}$$

Thus the distribution-free hit bounds are 0.100, 0.050, and 0.0192 for $k = 2, 3, 5$, respectively. Table 2 compares these numbers with exact probabilities for Gaussian, Laplace, and Student-t laws, and with a numerical quadrature calculation for a symmetric NIG law with $\alpha = \delta = 1$, $\beta = \mu = 0$. The exact derivations are collected in Section V.

	$k=2$		$k=3$		$k=5$	
Law	exact	bound/exact	exact	bound/exact	exact	bound/exact
Gaussian	0.02520	3.97	0.01102	4.54	0.00225	8.55
Laplace	0.02489	4.02	0.00916	5.46	0.00124	15.52
Cauchy / t_1	0.02598	3.85	0.01048	4.77	0.00171	11.26
Student t_3	0.02552	3.92	0.01100	4.54	0.00213	9.05
Student t_5	0.02533	3.95	0.01102	4.54	0.00220	8.73
Student t_{10}	0.02523	3.96	0.01102	4.54	0.00224	8.59
NIG, symmetric	0.02615	3.82	0.01062	4.71	0.00172	11.18

Table 2. Exact high-information loss probabilities versus the sharp distribution-free Cantelli bound (79). The ratio reports the universal bound divided by the exact probability. The NIG row is evaluated by one-dimensional numerical quadrature and root finding for the symmetric parameter choice $\alpha = \delta = 1$, $\beta = \mu = 0$.

The table has two messages. First, around moderate score thresholds $k = 2$ or $k = 3$, the sharp distribution-free envelope is typically about four or five times the exact hit probability in these classical laws. This is conservative, but not vacuous. Second, the bound becomes looser in the far score tail. This is unavoidable for a two-moment distribution-free inequality: Cantelli decays quadratically in a, while Gaussian, Laplace, Student-t self-information tails, and NIG self-information tails have exponential or faster decay in the information coordinate.

For the joint profile, the practical distribution-free bound

$$K_-(u, h_- + a) \le \min\left\{ \bar{F}_-(u),\ p_- \frac{V_-}{V_- + a^2} \right\} \tag{80}$$

behaves especially well in monotone symmetric examples. In such laws, magnitude and self-information thresholds are nested, so the exact joint probability is the smaller of the magnitude tail and the exact information-hit probability. The only looseness in (80) is therefore the Cantelli looseness displayed in Table 2; if the magnitude threshold binds, the joint-profile bound is exact.

The same comparison also clarifies how to use the path certificate. The simpler union-set bound in Theorem III.15 is conservative because the body bound $\widehat{B}_T(c, u)$ removes large raw losses but does not exploit the score cutoff. When log-scores are available, the score-filtered refinement (53) should be used: raising a lowers the hit budget but admits more high-information loss states back into the body term, while lowering a does the reverse. This creates the intended body-versus-hit tradeoff without estimating a pointwise density level or a tail index.

III-K General regime readings and the Donsker boundary

The set-indexed theorem supplies a single regime language:

finite-horizon path concentration = body accumulation while avoiding A + adverse-set hit risk.

Adverse-hit dominance means $H_T(A)$ is large relative to $B_T(c;A)$; body dominance means the body term is the main contributor. Choosing $A = A_u$ gives the usual magnitude-tail large-step reading. Choosing $A = A^{\cup}_{u,a}$ gives the practical magnitude-or-information certificate. Choosing a pure or joint information set expresses the same comparison through $K_-(t)$ or $K_-(u,t)$.

The finite-variance central-limit theorem is a boundary condition for this comparison in the asymptotic limit [4]. The following profile records whether one-step losses at the Brownian $\sqrt{T}$ scale are already negligible at the investor's actual horizon.

Proposition III.29 (Donsker boundary condition)**.** *Assume $\mathbb{E}X = 0$ and $0 < \sigma^2 = \mathbb{E}X^2 < \infty$. For $\varepsilon > 0$, define*

$$L_-(T,\varepsilon) := \frac{1}{\sigma^2}\mathbb{E}\big[X^2 \mathbf{1}_{\{X \le -\varepsilon\sigma\sqrt{T}\}}\big].$$

Then

$$T\bar{F}_-(\varepsilon\sigma\sqrt{T}) \le \varepsilon^{-2} L_-(T,\varepsilon). \tag{81}$$

The right-hand side tends to zero as $T \to \infty$. Thus a small $L_-(T,\varepsilon)$ certifies that one-step losses beyond the $\varepsilon\sigma\sqrt{T}$ scale are negligible. A non-small value is not a lower bound, but it marks a pre-asymptotic region in which Donsker's limiting theorem alone does not justify discarding the adverse-set term.

Proof. By Markov's inequality on the truncated second moment,

$$\bar{F}_-(\varepsilon\sigma\sqrt{T}) \le \frac{\mathbb{E}[X^2 \mathbf{1}_{\{X \le -\varepsilon\sigma\sqrt{T}\}}]}{\varepsilon^2\sigma^2 T}.$$

Multiplying by T gives (81). The convergence $L_-(T,\varepsilon) \to 0$ follows from dominated convergence. □

Remark III.30. Mean drawdown is a bulk average over the whole path distribution and can be dominated by central-mass effects. The mandate bounded in this paper is a floor-breach event, for which the set-indexed certificate separates ordinary body accumulation from adverse-set hit risk.

III-L A second-order information statement: the dispersion role of downside varentropy

The Cantelli certificate of Section III-E uses only the first two loss-side information moments and is, by Theorem III.12, the sharpest distribution-free bound available from them. This subsection records what the same two moments deliver *with* a regularity assumption, namely a second-order (dispersion) refinement of the loss-side information content. The result is the finite-sample, Berry–Esseen form of the statement that downside varentropy is the dispersion term of the asymptotic equipartition property restricted to the loss side. We are deliberately precise about what it does and does not assert: it characterizes the fluctuations of the *cumulative loss-side information content*, not the dispersion of the running minimum $\underline{S}_T$ itself, which is governed by the return law rather than by the self-information law.

Throughout, condition on the loss side. Let $Y := \imath(X) \mid \{X < 0\}$ have mean h_-, variance $V_- \in (0,\infty)$, and finite third absolute central moment

$$\rho_- := \mathbb{E}\Big[\,|\,\imath(X) - h_-\,|^3 \;\Big|\; X < 0\Big] < \infty.$$

Let $Y_1,\dots,Y_m$ be i.i.d. copies of Y (the information scores of the m loss-side increments in a sample) and write $\bar{S}_m := \sum_{i=1}^{m}(Y_i - h_-)$ for the centered cumulative loss-side information content.

Theorem III.31 (Second-order loss-side AEP with explicit rate)**.** *With the notation above,*

$$\sup_{r\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\bar{S}_m}{\sqrt{m\,V_-}} \le r\right) - \Phi(r) \right| \;\le\; \frac{C_0\,\rho_-}{V_-^{3/2}\sqrt{m}}, \qquad C_0 \le 0.4748. \tag{82}$$

Consequently the loss-side information content of m increments concentrates as

$$\sum_{i=1}^{m} \imath(X_i) \;=\; m\,h_- \;+\; \sqrt{m\,V_-}\,\zeta_m, \qquad \zeta_m \Rightarrow \mathcal{N}(0,1),$$

with the uniform error (82)*. In particular V_- is the dispersion in the loss-side AEP, and the standardized excess information $\big(\sum_i \imath(X_i) - m h_-\big)/\sqrt{m\,V_-}$ is asymptotically pivotal at rate $m^{-1/2}$.*

Proof. The summands $Y_i - h_-$ are i.i.d. with mean 0, variance V_-, and third absolute moment $\rho_- < \infty$. Inequality (82) is the classical Berry–Esseen theorem applied to this triangular-free i.i.d. sum; the uniform constant $C_0 \le 0.4748$ is the best known value [11]. The displayed concentration statement is the restatement of (82) after multiplying through by $\sqrt{mV_-}$ and adding mh_-. □

The next corollary is the bridge to the hit budget. It contrasts the distribution-free Cantelli decay (quadratic in the threshold) with the Gaussian decay available once the dispersion regime is entered, and quantifies the gap the calibration table reports numerically.

Corollary III.32 (Two regimes for the high-information hit budget)**.** *Fix an excess-information threshold of the form $a = k\sqrt{V_-}$, $k > 0$.*

1. (Distribution-free.) *Without any regularity assumption, Theorem III.11 gives*

$$\mathbb{P}\{\imath(X) - h_- \ge a \mid X < 0\} \;\le\; \frac{1}{1+k^2}.$$

2. (Block dispersion, fixed k.) *Let Y_i be the loss-side information scores in Theorem III.31. Then for the block excess threshold $k\sqrt{mV_-}$,*

$$\left|\mathbb{P}\left\{\sum_{i=1}^{m}(Y_i - h_-) \ge k\sqrt{mV_-}\right\} - \Phi(-k)\right| \le \frac{C_0\rho_-}{V_-^{3/2}\sqrt{m}}.$$

   In particular, for each fixed k, the block-information exceedance probability is $\Phi(-k) + O(m^{-1/2})$.

If the standardized score is exactly Gaussian, then the Gaussian value $\Phi(-k)$ is exact and the ratio between the Cantelli bound and the Gaussian value diverges as $k \to \infty$. Without exact Gaussianity, or without an additional moderate-deviation theorem, the Berry–Esseen result above is an additive fixed-k approximation. It should not be read as a relative far-tail approximation as k grows.

Proof. Part (1) is (79) with p_- divided out, i.e. $V_-/(V_-+a^2)$ at $a = k\sqrt{V_-}$. For part (2), apply Theorem III.31 and take complements. The Berry–Esseen inequality is uniform in the threshold, so the same additive error controls the upper tail at the fixed point k. The final statement follows from the Gaussian tail value in the exact-Gaussian case and from Mills' ratio; the caution follows because Berry–Esseen gives an additive error, not a relative tail expansion. □

Remark III.33 (What Theorem III.31 does and does not say). The theorem is a statement about the *information content* $\sum_i \imath(X_i)$, whose first two cumulants are mh_- and mV_-. It is *not* a dispersion statement for the wealth path $\underline{S}_T$: the second-order term of a finite-horizon breach probability is governed by the tilted variance of the *return* X, not by the variance of the self-information $\imath(X)$. The role of Theorem III.31 is to justify, at an explicit rate, the use of V_- (rather than only the loss frequency p_-) as the operative finite-horizon summary of *informational* surprise: it is precisely the quantity whose square root sets the scale of cumulative-surprisal fluctuations. The certificate then converts that surprise into a breach budget through the Cantelli witness, which holds distribution-free and does not require the dispersion regime of (82).

IV Economic consequences and practitioner use

The preceding section produced a finite-horizon floor-breach certificate from estimable loss-side quantities. This section turns that certificate into a tool for choosing portfolios, and asks what it adds to the controls a practitioner already uses. Three questions organize the section. First, how does the body/hit budget enter portfolio construction as an explicit constraint, rather than as a post-hoc leverage adjustment section IV-A? Second, can the constraint be imposed in a multi-asset allocation without sacrificing convexity, given that the exact path-concentration feasible set is itself nonconvex (section IV-B)? Third—and this is the question on which the section turns—what does path-concentration control actually contribute beyond a variance target or a Busseti–Ryu–Boyd drawdown surrogate, once one accounts for the fact that all three are functionals of the same adverse loss mass (section IV-D)?

The answer to the third question is deliberately not a portfolio horse race. Path-concentration control and an aggressive negative-power BRB certificate target overlapping adverse mass, so the real comparison question is not which one wins but where they differ: the certificate is built from an integrated, well-estimated information functional, whereas a strong BRB moment requires estimating a near-maximum that is unstable in exactly the finite samples where tail risk matters. Section IV-D isolates this contrast on a matched pair of synthetic laws and identifies the regime in which the two controls agree—the method's intended null—as carefully as the regime in which they separate.

Let $A \subseteq (-\infty, 0)$ be the adverse set and let

$$B_T(c; A) = \mathbb{P}\{\underline{S}_T \le -c,\ X_i \notin A \text{ for all } i \le T\}$$

be the body breach probability on paths avoiding A. The hit probability is

$$H_T(A) = 1 - (1 - p_A)^T, \qquad p_A = \mathbb{P}(X \in A).$$

For a certified body bound $\overline{B}_T(c; A) \ge B_T(c; A)$, the master certificate is

$$\mathbb{P}(\underline{S}_T \le -c) \le \overline{B}_T(c; A) + H_T(A) \le \overline{B}_T(c; A) + T p_A. \tag{83}$$

The baseline distribution-free implementation uses the union adverse set

$$A^{\cup}_{u,a} = \{X \le -u\} \cup \{X < 0 : \iota(X) - h_- \ge a\},$$

whose hit probability is bounded by

$$r_C(u, a) = \left(\bar{F}_-(u) + p_- \frac{V_-}{V_- + a^2}\right) \wedge p_-.$$

The resulting certificate is

$$\mathbb{P}(\underline{S}_T \le -c) \le \hat{B}_T(c, u) + 1 - (1 - r_C(u, a))^T, \tag{84}$$

with the score-filtered refinement (53) used when loss-side score observations are available. The terms in (84) are either empirical frequencies, sample averages, or integrated information functionals. In particular, the baseline certificate does not require a tail index, a parametric family, a pointwise estimate of $\iota(-u)$, or a fitted predictive density at the threshold.

IV-A Path-concentration budgets for portfolio construction

We believe that the most direct use case is in portfolio construction, which should not be confused with a post-hoc leverage throttle. A practitioner may already compute a growth-optimal Kelly portfolio, or may solve the Busseti–Ryu–Boyd risk-constrained Kelly problem, which adds a drawdown-probability constraint to the classical Kelly problem and replaces the hard constraint by a tractable bound leading to a convex optimization problem [3]. Our contribution is to add a separate finite-horizon path-concentration budget: choose portfolios whose floor breach certificate is not dominated by adverse high-information hits.

The useful implementation splits the path budget into a body budget and an adverse-hit budget. For a total path-concentration budget α_{pc} and a chosen hit share $s_0 \in (0, 1)$, set

$$\alpha_{\text{hit}} = s_0 \alpha_{\text{pc}}, \qquad \alpha_{\text{body}} = (1 - s_0)\alpha_{\text{pc}}.$$

A portfolio is admissible if its certified body bound and certified adverse-hit bound satisfy

$$\overline{B}_T(c; A) \le \alpha_{\text{body}}, \qquad H_T(A) \le \alpha_{\text{hit}}. \tag{85}$$

The parameter s_0 is not a timing gate. It is a risk-budget design choice: small s_0 is strict about rare adverse-hit exposure, while large s_0 allows more of the certified risk to come from adverse cells.

The resulting optimization asks a different question from BRB. BRB controls a drawdown-probability surrogate. The path-concentration constraint controls the composition of that surrogate, separating ordinary body accumulation from rare adverse-hit exposure. In particular, two portfolios can have similar expected log growth, volatility, or BRB score while having very different values of $\overline{B}_T$ and H_T.

It is important not to overstate the distinction between BRB and PC, since they are functionals of largely the same adverse loss mass. At the *population* level, PC need not add allocative value against *a correctly specified* and *aggressive* negative-power certificate. The BRB moment

$$M_\lambda = \mathbb{E}e^{-\lambda X}$$

is already an exponential functional of the same adverse loss mass, and for large λ it makes those losses expensive. At the level of *finite samples*, however, when estimation is required, there is a very clear distinction, since strong negative-power certificate controls require estimating a near-max functional that is often unreliable. PC controls that mass through an integrated, well-estimated functional, is robust across the BRB risk-aversion exponent rather than requiring it to be chosen correctly, and — independently of the allocation question — resolves finite-horizon risk into body vs. hit components that any scalar drawdown surrogate discards.

IV-B Multi-asset allocation and convex certificates

The same overlay can be used in a multi-asset Kelly or BRB allocation problem, but the exact path-concentration constraint should not be expected to be convex. Let $R \in \mathbb{R}^d$ be the one-period simple-return vector and let $w \in \mathbb{R}^d$ be portfolio weights. The one-period portfolio log return is

$$X_w = \log(1 + w^\top R),$$

defined on the solvency domain $1 + w^\top R > 0$. The exact finite-horizon constraint is

$$\mathbb{P}\left(\min_{0 \le k \le T} \sum_{j=1}^{k} X_{w,j} \le -c \right) \le \alpha_{\text{pc}}. \tag{86}$$

This is a chance constraint for a path functional. It is economically meaningful, but it is not generally convex in w.

Proposition IV.1 (The exact path-concentration feasible set need not be convex)**.** *Even for $T = 1$, the feasible set associated with (86) can be nonconvex.*

Proof. Take two assets and two equiprobable scenarios

$$R^{(1)} = (-0.6, 0), \qquad R^{(2)} = (0, -0.6).$$

Set the floor $c = u = -\log(3/4)$, so that $X_w \le -c$ is equivalent to $1 + w^\top R \le 3/4$. With $\alpha_{\text{pc}} = 1/2$, the portfolios $w^{(1)} = (1, 0)$ and $w^{(2)} = (0, 1)$ are feasible: each breaches the one-period floor in exactly one of the two scenarios. Their average $\bar{w} = (1/2, 1/2)$, however, has gross return $1 + \bar{w}^\top R = 0.7 < 3/4$ in both scenarios, so the breach probability is one. Thus two feasible portfolios can have an infeasible average. □

The practical analogue of the BRB philosophy is therefore to replace the exact chance constraint by a conservative convex certificate. For $q > 0$ define the negative-power moment

$$M_q(w) := \mathbb{E}\left[(1 + w^\top R)^{-q}\right]. \tag{87}$$

The next result gives the convex certificate used by the multi-asset overlay. A score envelope $\Phi(a)$ is treated as fixed at the optimization stage; it may be the Pareto-score envelope, or a distribution-free Cantelli envelope using an a priori uniform bound on downside varentropy over the admissible portfolio class. Portfolio-specific estimates of $V_{-}(w)$ can still be used for post-optimization reporting or in an outer-loop recalibration, but they are not assumed to be convex functions of w.

Theorem IV.2 (Convex path-concentration certificate)**.** *Fix $T \in \mathbb{N}$, $c > 0$, $u > 0$, $a > 0$, and $\theta, \eta > 0$. Suppose that for all portfolios under consideration*

$$\mathbb{P}\{X_w < 0,\ \iota_w(X_w) - h_{-,w} \ge a \mid X_w < 0\} \le \Phi(a),$$

where $\Phi(a) \in [0, 1]$ is fixed. Then the following condition is sufficient for (86):

$$e^{-\theta c} \sum_{k=1}^{T} M_\theta(w)^k + T\, M_\eta(w)\, (e^{-\eta u} + \Phi(a)) \le \alpha_{\text{pc}}. \tag{88}$$

Moreover, on any convex set on which $1 + w^\top R > 0$ almost surely and the moments in (87) are finite, the left-hand side of (88) is a convex function of w.

Proof. Let

$$A_w = \{X_w \le -u\} \cup \{X_w < 0,\ \iota_w(X_w) - h_{-,w} \ge a\}.$$

The set-indexed theorem gives

$$\mathbb{P}(\underline{S}_T(w) \le -c) \le B_T(c; A_w) + T\mathbb{P}(X_w \in A_w),$$

where $\underline{S}_T(w) = \min_{0 \le k \le T} \sum_{j=1}^{k} X_{w,j}$. Since $A_w \supseteq \{X_w \le -u\}$, avoiding A_w implies $X_w > -u$. Therefore, for any $\theta > 0$, Markov's inequality and a union bound over crossing times give

$$B_T(c; A_w) \le e^{-\theta c} \sum_{k=1}^{T} \mathbb{E}\left[e^{-\theta X_w} \mathbf{1}_{\{X_w > -u\}}\right]^k \le e^{-\theta c} \sum_{k=1}^{T} M_\theta(w)^k.$$

For the hit probability, Markov's inequality gives

$$\mathbb{P}(X_w \le -u) = \mathbb{P}(e^{-\eta X_w} \ge e^{\eta u}) \le e^{-\eta u} M_\eta(w),$$

and also

$$\mathbb{P}(X_w < 0) \le \mathbb{P}(e^{-\eta X_w} > 1) \le M_\eta(w).$$

The assumed score envelope implies

$$\mathbb{P}\{X_w < 0,\ \iota_w(X_w) - h_{-,w} \ge a\} \le \mathbb{P}(X_w < 0)\Phi(a) \le M_\eta(w)\Phi(a).$$

Thus

$$\mathbb{P}(X_w \in A_w) \le M_\eta(w)(e^{-\eta u} + \Phi(a)).$$

Combining these estimates proves sufficiency of (88).

It remains to prove convexity. For fixed $q > 0$, the map $z \mapsto z^{-q}$ is convex on $(0, \infty)$. Since $w \mapsto 1 + w^\top R$ is affine, $(1 + w^\top R)^{-q}$ is convex in w on the solvency domain. Taking expectations preserves convexity, so $M_q(w)$ is convex and nonnegative. For each integer $k \ge 1$, the function $x \mapsto x^k$ is convex and nondecreasing on $[0, \infty)$, hence $M_\theta(w)^k$ is convex by composition. The first term in (88) is therefore a nonnegative sum of convex functions, and the second is a nonnegative multiple of the convex function $M_\eta(w)$. This proves convexity. □

Two implementation details are worth making explicit. First, the certificate keeps θ, η, u, and a fixed inside the convex problem. One can solve a grid of convex problems over these parameters and choose the best certificate afterward; placing an infimum over them inside the constraint would not, in general, preserve convexity. Second, (88) uses the rare-hit upper bound $H_T(A_w) \le T\mathbb{P}(X_w \in A_w)$. The exact expression $1 - (1 - p)^T$ is tighter but concave in p, so it is not the natural object for convex certification.

A convex multi-asset overlay can therefore be written as

$$\begin{aligned} \max_{w} \quad & \mathbb{E}\left[\log(1 + w^\top R)\right] \\ \text{s.t.} \quad & C_{\mathrm{BRB}}(w) \le 0, \\ & e^{-\theta c} \sum_{k=1}^{T} M_\theta(w)^k + T\, M_\eta(w)\,(e^{-\eta u} + \Phi(a)) \le \alpha_{\mathrm{pc}}, \\ & w \in \mathcal{C}, \end{aligned} \tag{89}$$

where $\mathcal{C}$ contains ordinary portfolio constraints and the solvency condition. If $\mathcal{C}$ is convex and C_{BRB} is a convex BRB certificate, then (89) is a convex optimization problem in the standard sense: maximize a concave objective over a convex feasible set. Equivalently, one may minimize the convex function $-\mathbb{E}[\log(1 + w^\top R)]$.

The controlled experiment below does not solve (89). It isolates the certificate's discrimination problem in a two-law setting where population quantities and ground-truth floor-breach probabilities are known. The convex program is a scalability result for multi-asset implementation, not the object being numerically optimized in the synthetic mechanism test.

IV-C A minimax separation between the BRB input and the PC certificate

Fix a finite tilt grid $\Theta = \{\theta_1 < \cdots < \theta_m\} \subset (0, \infty)$, $\theta_{\max} = \theta_m$, $\theta_{\min} = \theta_1$. The grid body bound and the population certificate are

$$B_T^\Theta(c, u) := \min_{\theta \in \Theta} e^{-\theta c} \sum_{k=1}^{T} m_u(\theta)^k, \qquad \mathrm{PC}^\Theta(P) := B_T^\Theta(c, u) + T\, r_C(u, a). \tag{90}$$

Since the grid minimum dominates the infimum, $B_T^\Theta(c, u) \ge \widehat{B}_T(c, u)$, so the linearized certificate (42) gives

$$\mathbb{P}(\underline{S}_T \le -c) \le \mathrm{PC}^\Theta(P). \tag{91}$$

For i.i.d. $X_1, \dots, X_n$ the plug-in certificate is

$$\widehat{\mathrm{PC}}_n := \min_{\theta \in \Theta} e^{-\theta c} \sum_{k=1}^{T} \widehat{m}_n(\theta)^k + T\, \widehat{r}_{C,n}, \tag{92}$$

with $\widehat{p}_{-,n} = \frac{1}{n}\sum_i \mathbf{1}_{\{X_i < 0\}}$, $\widehat{F}_n(u) = \frac{1}{n}\sum_i \mathbf{1}_{\{X_i \le -u\}}$, $\widehat{m}_n(\theta) = \frac{1}{n}\sum_i e^{-\theta X_i}\mathbf{1}_{\{X_i > -u\}}$, $\widehat{V}_{-,n}$ the loss-side score sample variance ($(n_- - 1)$-normalized) when $n_- \ge 2$, and set $\widehat{V}_{-,n} = 0$ when $n_- < 2$, and $\widehat{r}_{C,n} = \left(\widehat{F}_n(u) + \widehat{p}_{-,n}\widehat{V}_{-,n}/(\widehat{V}_{-,n} + a^2)\right) \wedge \widehat{p}_{-,n}$.

Remark IV.3 (Oracle status of the minimax PC statement). The minimax statement in this subsection is an oracle statistical statement. It assumes that the loss-side information scores

$$Y_i = \iota(X_i) \mid \{X_i < 0\}$$

are observed for the loss observations, so that V_- is estimated from score observations. This is different from the fully nonparametric implementation in Section IV-D, where the scores are replaced by spacing or nearest-neighbor proxies. Thus

$$\text{oracle PC certificate} \quad \not\equiv \quad \text{kNN-estimated PC certificate.}$$

The oracle theorem isolates the statistical difficulty of the PC inputs once scores are available. Consistency, bias, and concentration of a particular score estimator, such as the k-nearest-neighbor estimator used below, are separate implementation questions.

Definition IV.4 (Tail-location-free class). Fix $u > 0$ and constants $p_{\min} \in (0, 1]$, $0 < V_{\min} \le V_{\max} < \infty$, $\kappa_4 < \infty$. Let $\mathcal{P}(u) = \mathcal{P}(u; p_{\min}, V_{\min}, V_{\max}, \kappa_4)$ be the set of absolutely continuous laws on $\mathbb{R}$ whose loss side satisfies

$$p_- \in [p_{\min}, 1], \qquad V_- \in [V_{\min}, V_{\max}], \qquad \mathbb{E}\big[(\iota(X) - h_-)^4 \mid X < 0\big] \le \kappa_4,$$

with *no* constraint on the location of loss-side mass. The class is *nondegenerate* if $V_{\min} < V_{\max}$ and κ_4 strictly exceeds the fourth central score moment of some two-shelf loss law supported in $(-u, 0)$ with loss mass at least $p_{\min}$ and score variance interior to $(V_{\min}, V_{\max})$.

Theorem IV.5 (Minimax separation: BRB input versus PC certificate). *Let $\mathcal{P}(u)$ be nondegenerate, $T \ge 1$, $0 < u \le c$, and $a > 0$.*

(i) (The aggressive-BRB input is minimax-inestimable.) *For every $\lambda > 0$ and every sample size $n \ge 2$,*

$$\inf_{\widehat{M}} \sup_{P \in \mathcal{P}(u)} \mathbb{E}_P \Big| \widehat{M} - M_\lambda(P) \Big| = \infty,$$

the infimum taken over all estimators $\widehat{M} = \widehat{M}(X_1, \dots, X_n)$. If $\mathcal{P}(u)$ is further restricted to losses bounded below by $-M$, the same two-point construction gives the lower bound

$$\inf_{\widehat{M}} \sup_{P \in \mathcal{P}(u)} \mathbb{E}_P |\widehat{M} - M_\lambda(P)| \gtrsim \frac{e^{\lambda M}}{n}.$$

We do not claim this is the sharp minimax rate on the bounded-loss subclass.

(ii) (Every PC input is minimax-estimable at the parametric rate.) *There is $C = C(T, c, u, a, \Theta, p_{\min}, V_{\min}, V_{\max}, \kappa_4)$ such that, when the loss-side scores are observed (oracle plug-in for V_-),*

$$\sup_{P \in \mathcal{P}(u)} \mathbb{E}_P \Big| \widehat{\mathrm{PC}}_n - \mathrm{PC}^{\Theta}(P) \Big| \le C\, n^{-1/2}.$$

(iii) (One-sided oracle certification under fourth moments.) For every $\gamma \in (0, 1/2)$ and every sample size

$$n \ge n_\gamma := \left\lceil \frac{8}{p_{\min}} \log \frac{3}{\gamma} \vee \frac{4}{p_{\min}} \right\rceil,$$

there is a margin

$$\Delta_n(\gamma) = C' \left\{ \sqrt{\frac{\log(|\Theta|/\gamma)}{n}} + \sqrt{\frac{1}{n p_{\min} \gamma}} \right\},$$

where $C' = C'(T, c, u, a, \Theta, p_{\min}, V_{\min}, V_{\max}, \kappa_4)$, such that the oracle upper certificate

$$\mathrm{PC}_n^{\mathrm{or}} := \widehat{\mathrm{PC}}_n + \Delta_n(\gamma)$$

satisfies

$$\inf_{P \in \mathcal{P}(u)} P\big\{ P(\underline{S}_T \le -c) \le \mathrm{PC}_n^{\mathrm{or}} \big\} \ge 1 - \gamma.$$

Equivalently, under only the fourth-moment score assumption, the uniform one-sided margin is $O((n\gamma)^{-1/2})$, up to the logarithmic bounded-mean terms. A sharper $O(n^{-1/2}\sqrt{\log(1/\gamma)})$ margin requires additional boundedness or sub-exponential assumptions on the loss-side scores, or a robust variance estimator with the corresponding concentration theorem.

Proof. (i). Fix $\lambda > 0$ and $n \geq 2$. By nondegeneracy choose a two-shelf loss law $P_0 \in \mathcal{P}(u)$ whose loss side is supported in $(-u, 0)$, so $P_0(X \leq -u) = 0$, with $V_-(P_0)$ interior to $(V_{\min}, V_{\max})$ and $\mu_4(P_0) < \kappa_4$ strictly; write $p_0 := p_-(P_0) \geq p_{\min}$. Fix a width $\varepsilon_0 > 0$. For $0 < \delta < 1$ and $M > u + \varepsilon_0/2$, let

$$U_M = \mathrm{Unif}\big[-M - \varepsilon_0/2,\, -M + \varepsilon_0/2\big], \qquad P_{\delta,M} := (1-\delta)P_0 + \delta U_M.$$

The far shelf has self-information $\log(\varepsilon_0/\delta)$, independent of M. Hence the conditional loss-side score law of $P_{\delta,M}$ does not depend on M. Its first four centered score moments differ from those of P_0 by at most $O(\delta \log^4(1/\delta))$, which tends to zero as $\delta \downarrow 0$. Since $V_-(P_0)$ and $\mu_4(P_0)$ are strictly interior, there exists $\bar\delta > 0$, depending only on the class constants and on how interior P_0 is, such that

$$P_{\delta,M} \in \mathcal{P}(u) \qquad \text{for all } 0 < \delta \leq \bar\delta \text{ and all } M > u + \varepsilon_0/2.$$

Also

$$p_-(P_{\delta,M}) = (1-\delta)p_0 + \delta \geq p_0 \geq p_{\min}.$$

For the given sample size $n \geq 2$, set

$$\delta_n := \min\{1/n, \bar\delta\}, \qquad P_M := P_{\delta_n, M}.$$

Then $\delta_n \leq 1/n$, and

$$c_0 := \inf_{m\geq 2} m\delta_m = \inf_{m \geq 2} \min\{1, m\bar\delta\} = \min\{1, 2\bar\delta\} > 0.$$

Thus $\delta_n \geq c_0/n$ for every $n \geq 2$.

Since P_0 and U_M have disjoint supports, $dP_0/dP_M = (1-\delta_n)^{-1}$ on $\mathrm{supp}(P_0)$, while P_0 charges none of U_M's support. Therefore

$$\mathrm{KL}(P_0 \| P_M) = \log \frac{1}{1-\delta_n}, \qquad \mathrm{KL}\big(P_0^{\otimes n} \| P_M^{\otimes n}\big) = n \log \frac{1}{1-\delta_n} \leq n \log \frac{1}{1 - 1/n} \leq 2 \log 2.$$

The two-point Le Cam bound [12, Thm. 2.2], applied to the functional M_λ with gap

$$\Delta_M := M_\lambda(P_M) - M_\lambda(P_0),$$

gives

$$\inf_{\widehat{M}} \max_{P \in \{P_0, P_M\}} E_P\Big|\widehat{M} - M_\lambda(P)\Big| \geq \frac{\Delta_M}{4} \exp\big\{-\mathrm{KL}\big(P_0^{\otimes n} \| P_M^{\otimes n}\big)\big\} \geq \frac{1}{16}\Delta_M.$$

Finally,

$$\Delta_M = \delta_n\Big(E_{U_M}[e^{-\lambda X}] - M_\lambda(P_0)\Big) \geq \frac{c_0}{n}\left(e^{\lambda(M-\varepsilon_0/2)} - M_\lambda(P_0)\right) \to \infty \qquad \text{as } M \to \infty.$$

Taking the supremum over $M > u + \varepsilon_0/2$ proves the divergence.

If the class is additionally restricted to losses bounded below by $-L$, use instead a far shelf

$$U_L = \mathrm{Unif}[-L,\, -L + \varepsilon_0],$$

with $L > u + \varepsilon_0$. The same argument gives

$$\Delta_L \geq \frac{c_0}{n}\left(e^{\lambda(L-\varepsilon_0)} - M_\lambda(P_0)\right),$$

so the two-point construction yields the lower bound

$$\inf_{\widehat{M}} \sup_{P \in \mathcal{P}(u)} E_P|\widehat{M} - M_\lambda(P)| \gtrsim \frac{e^{\lambda L}}{n}$$

on the $X \geq -L$ bounded-loss subclass. Relabeling L as M gives the bounded-loss lower bound stated in part (i).

(ii). Each input is a bounded empirical mean or a controlled variance, uniformly over $\mathcal{P}(u)$. The counts obey $\mathbb{E}|\widehat{p}_{-,n} - p_-| \leq \frac{1}{2\sqrt{n}}$ and $\mathbb{E}|\widehat{F}_n(u) - \bar{F}_-(u)| \leq \frac{1}{2\sqrt{n}}$ with no condition, being means of $\{0,1\}$ variables; crucially $\widehat{F}_n(u)$ depends only on whether each loss exceeds u in magnitude, not on how far. For the truncated transform, $e^{-\theta X}\mathbf{1}_{\{X > -u\}} \in [0, e^{\theta u}]$, so $\mathbb{E}|\widehat{m}_n(\theta) - m_u(\theta)| \leq e^{\theta u}/(2\sqrt{n}) \leq e^{\theta_{\max} u}/(2\sqrt{n})$. The map $(m(\theta))_{\theta\in\Theta} \mapsto \min_\theta e^{-\theta c}\sum_{k=1}^T m(\theta)^k$ is Lipschitz on $\prod_\theta [0, e^{\theta u}]$ with constant $L_B := e^{-\theta_{\min} c}\sum_{k=1}^T k\, e^{(k-1)\theta_{\max} u}$ (the minimum is 1-Lipschitz in the sup norm, and each summand has m-derivative $\leq k\, e^{(k-1)\theta_{\max} u}$), so the empirical grid body bound has error at most $L_B\, e^{\theta_{\max} u}/(2\sqrt{n})$. For the score variance, conditional on $n_- = \sum_i \mathbf{1}_{\{X_i < 0\}} \geq 2$ observed scores the unbiased sample variance satisfies $\mathbb{E}[(\widehat{V}_{-,n} - V_-)^2 \mid n_-] \leq \kappa_4/n_-$, hence $\mathbb{E}|\widehat{V}_{-,n} - V_-| \leq \sqrt{\kappa_4}\,\mathbb{E}[n_-^{-1/2}]$; since $n_- \sim \mathrm{Bin}(n, p_-)$ with $p_- \geq p_{\min}$, a Chernoff bound gives $\mathbb{E}[n_-^{-1/2}] \leq C'(p_{\min})\, n^{-1/2}$ for $n \geq n_1(p_{\min})$. Finally $g(p, F, V) = (F + pV/(V + a^2)) \wedge p$ is Lipschitz on $[0,1]^2 \times$

$[V_{\min}, V_{\max}]$, with gradient bounded by $\max\{2,\ a^2/(V_{\min}+a^2)^2\}$, so $\mathbb{E}|\hat{r}_{C,n} - r_C(u,a)| \le C''(a, V_{\min})\, n^{-1/2}$. Combining the body and hit estimates yields $\mathbb{E}|\widehat{\mathrm{PC}}_n - \mathrm{PC}^{\Theta}(P)| \le C\, n^{-1/2}$ uniformly.

(iii). The bounded empirical means in the body part still have sub-Gaussian tails. Indeed, the count estimates $\hat{p}_{-,n}$, $\hat{F}_n(u)$, and the truncated transforms $\hat{m}_n(\theta)$ are empirical means of bounded variables, uniformly over $P \in \mathcal{P}(u)$. A Hoeffding bound, combined with a union bound over the finite grid Θ, gives deviations of order

$$O\left(\sqrt{\frac{\log(|\Theta|/\gamma)}{n}}\right)$$

for all bounded-mean inputs simultaneously.

The score-variance input is different. Under the present class assumption we only have a uniform fourth central moment for the loss-side score, so the ordinary sample variance need not have a Bernstein or Hoeffding tail. Let $n_- = \sum_{i=1}^n \mathbf{1}_{\{X_i<0\}}$.Since $p_- \ge p_{\min}$, a Chernoff bound gives

$$P(n_- < np_{\min}/2) \le \exp(-np_{\min}/8).$$

Under the sample-size condition in part (iii),

$$n \ge \frac{8}{p_{\min}} \log \frac{3}{\gamma},$$

this low-count probability is at most $\gamma/3$. The additional condition $n \ge 4/p_{\min}$ ensures that on the event $\{n_- \ge np_{\min}/2\}$ there are at least two loss-side score observations, so that the ordinary unbiased sample variance is well-defined. On the event $\{n_- \ge np_{\min}/2\}$, the unbiased sample variance of the oracle loss-side scores satisfies

$$E\left[(\hat{V}_{-,n} - V_-)^2 \mid n_-\right] \le \frac{\kappa_4}{n_-} \le \frac{2\kappa_4}{np_{\min}}.$$

Chebyshev's inequality therefore yields, on the good-count event,

$$P\left(|\hat{V}_{-,n} - V_-| > \sqrt{\frac{6\kappa_4}{np_{\min}\gamma}},\ n_- \ge np_{\min}/2\right) \le \frac{\gamma}{3}.$$

Combining this with the low-count event, whose probability is at most $\gamma/3$ by the sample-size condition, gives a total variance-input failure probability at most $2\gamma/3$. Enlarging the universal constant in $\Delta_n(\gamma)$, or equivalently allocating the bounded-mean deviations to the remaining probability budget, gives the stated $1-\gamma$ certificate.

The map

$$(p, F, V) \longmapsto \left(F + \frac{pV}{V + a^2}\right) \wedge p$$

is Lipschitz on $[0,1]^2 \times [0,\infty)$, with a constant depending only on a, and the grid body map is Lipschitz on the bounded transform range used in part (ii). Propagating the bounded-mean and score-variance deviations through these Lipschitz maps gives a constant C', depending only on the displayed class and certificate parameters, such that

$$PC^{\Theta}(P) \le \widehat{\mathrm{PC}}_n + C' \left\{ \sqrt{\frac{\log(|\Theta|/\gamma)}{n}} + \sqrt{\frac{1}{np_{\min}\gamma}} \right\}$$

with probability at least $1-\gamma$, uniformly over $P \in \mathcal{P}(u)$. Combining this inequality with the population validity

$$P(\underline{S}_T \le -c) \le PC^{\Theta}(P)$$

proves the oracle one-sided certification statement.

□

IV-D Controlled synthetic experiment: what PC adds beyond a scalar BRB moment

The experiment is a matched-pair simulation. We construct two synthetic one-period log-return laws, A and B, that agree in the summaries used by the competing controls—mean, variance, and a weak negative-power moment M_3—but differ in their loss-side magnitude-information profiles, and hence in genuine finite-horizon floor-breach risk. Holding the classical summaries fixed forces variance control and a weak BRB rule to be indifferent between the two laws by construction, so any separation that path-concentration control reports is attributable to the profile coordinate alone, not to a residual difference in level. We then vary the BRB exponent λ to locate the crossover at which a sufficiently aggressive negative-power certificate recovers the same separation, and report the finite-sample behavior of the two estimators in the regime where that crossover occurs.

Each one-period log-return law is a mixture of disjoint uniform shelves. If shelf j has center x_j, width ε_j, and probability mass p_j, then the density on that shelf is p_j/ε_j, the shelf score is

$$y_j = -\log(p_j/\varepsilon_j) = \log(\varepsilon_j/p_j),$$

and the negative-power moment is available in closed form:

$$M_\lambda = \sum_j p_j e^{-\lambda x_j} \frac{\sinh(\lambda\varepsilon_j/2)}{\lambda\varepsilon_j/2}.$$

The centers and masses determine mean, variance, and M_λ; the shelf widths determine the loss-side score dispersion. The construction is chosen so that Law B carries genuine breach-relevant adverse mass, not merely a different score width at the same magnitude law.

Table 3. Synthetic shelf laws used in the controlled experiment. Each row is a uniform shelf $[x_j - \varepsilon_j/2, x_j + \varepsilon_j/2]$ with mass p_j. Scores are $y_j = \log(\varepsilon_j/p_j)$. The positive variance-matching shelf in Law A is deliberately placed on the gain side so that A can match B's variance without adding an adverse rare-loss shelf of its own.

Law	Shelf role	Center	Mass	Width	Score
A	ordinary loss body	-1.20%	21.41%	0.60%	-3.57
A	ordinary gain body	0.40%	78.16%	0.20%	-5.97
A	positive variance-matching shelf	20.00%	0.43%	2.00%	1.54
B	rare adverse-hit shelf	-5.50%	3.00%	1.20%	-0.92
B	ordinary mild loss body	-0.40%	45.00%	0.10%	-6.11
B	ordinary gain body	0.40%	43.00%	0.20%	-5.37
B	moderate gain body	3.50%	9.00%	0.80%	-2.42

The two laws are matched in the classical quantities used by variance control and by the weak BRB moment M_3. The matching claim is intentionally narrow: the experiment matches M_3, not all weak or moderate negative-power moments. Law A has no one-period loss capable of breaching the -5% floor; its floor breaches are body-accumulation events. Law B has a breach-relevant adverse shelf: part of that shelf lies below -5%, and it has high loss-side score because a small probability mass is spread over a low-density adverse region.

Table 4. Population matched-pair diagnostics. The mandate is $T = 30$ periods and floor $c = 5\%$. Breach probabilities are Monte Carlo estimates from the specified population laws. The PC hit bound uses $u = 5\%$ and excess-score threshold $a = 3$. It is a conservative bound, not a calibrated probability forecast.

Law	Mean	Daily sd	M_3	V_-	$K_-(5\%, h_- + 3)$	PC hit bound	Terminal br.	Path br.
A: body-dominated	14.16 bp	1.46%	0.99659	0.00	0.00%	0.00%	3.53%	7.92%
B: adverse-hit	14.20 bp	1.46%	0.99672	1.58	2.75%	95.64%	12.22%	28.40%

Tables 4 and 5 give the main mechanism check. A variance rule is indifferent because the means and variances are essentially identical. A weak negative-power BRB rule at $\lambda = 3$ is also essentially indifferent because M_3 differs by only about 1.3×10^{-4}, or 0.013%. PC is not indifferent: the loss-side profile of Law B contains a nonzero joint adverse cell, $K_-(5\%, h_- + 3) = 2.75\%$, and its downside score dispersion is much larger.

The full decomposition in Table 5 shows why the mechanism is an adverse-hit mechanism rather than ordinary body accumulation. Law A has no adverse-set hit term at the chosen threshold, so its breaches are body events. Law B has very small body-breach probability after the adverse shelf is removed, but it has substantial adverse-set exposure. The certified hit term $\widehat{H}_T$ remains conservative, as expected from the distribution-free Cantelli envelope; its role here is to certify and rank the adverse-hit mechanism, not to provide a calibrated forecast of the exact path-breach probability.

Table 5. Full PC decomposition for the matched-pair experiment. Here $A = A^{\cup}_{u,a}$, with $T = 30$, $c = u = 5\%$, and $a = 3$. The body term $B_T(c; A)$ and the path-breach probability are Monte Carlo estimates under the specified population laws. The hit term $H_T(A) = 1 - (1 - p_A)^T$ uses the population adverse-set mass, while $\widehat{H}_T = 1 - (1 - r_C(u, a))^T$ is the Cantelli/PC hit certificate. The column $B_T + \widehat{H}_T$ reports the population-body plus certified-hit bound; in a purely empirical implementation B_T would be replaced by the computable body upper bound.

Law	$B_T(c; A)$	$H_T(A)$	$\widehat{H}_T$	$B_T + \widehat{H}_T$	Path br.
A: body-dominated	7.92%	0.00%	0.00%	7.92%	7.92%
B: adverse-hit	0.10%	59.90%	95.64%	95.74%	28.40%

The next diagnostic explains why this does not contradict BRB. As λ increases, the negative-power moment becomes exponentially more sensitive to the same adverse shelf that PC identifies. Thus a sufficiently aggressive BRB certificate eventually agrees with PC.

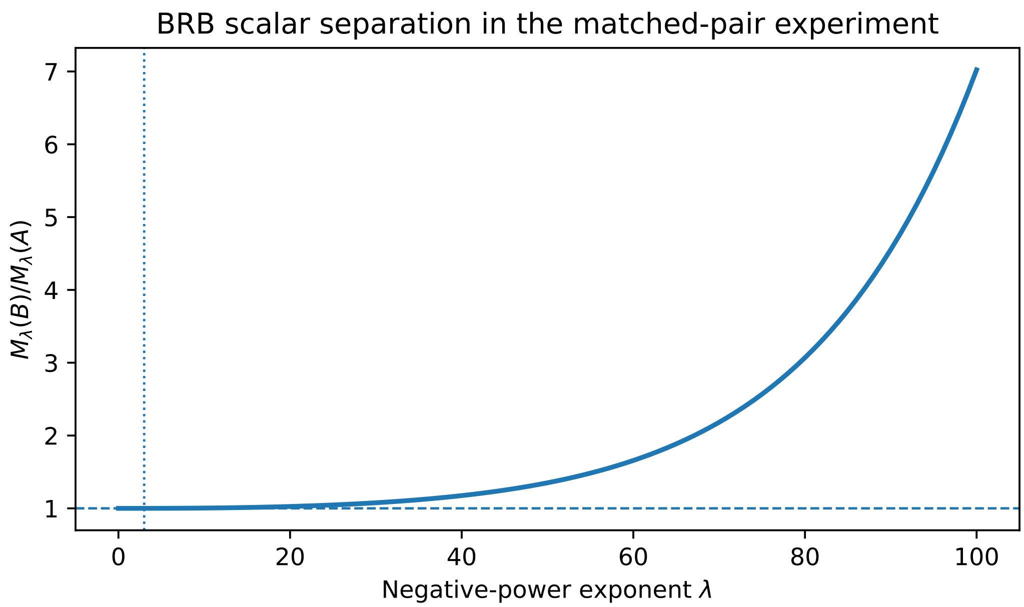

Figure 1. Scalar BRB separation in the synthetic matched-pair experiment. The weak moment M_3 is matched, so BRB is nearly indifferent at $\lambda = 3$. As λ grows, $M_\lambda(B)/M_\lambda(A)$ rises sharply, meaning an aggressive negative-power BRB certificate also identifies the adverse-hit law.

Table 6. Selected values from the BRB exponent sweep. The row $\lambda = 3$ is the near-matched weak-BRB point. Separation begins before the very large-λ regime; by $\lambda = 20$, $M_\lambda(B)$ is already 2.6% larger than $M_\lambda(A)$. Large λ values recover the adverse-cell separation because M_λ becomes an exponential loss functional.

λ	$M_\lambda(A)$	$M_\lambda(B)$	$M_\lambda(B)/M_\lambda(A)$
0	1.000	1.000	1.000
3	0.997	0.997	1.000
10	0.993	0.997	1.004
20	0.994	1.020	1.026
40	1.013	1.190	1.175
60	1.057	1.753	1.658
80	1.133	3.476	3.069
100	1.246	8.752	7.022

This is the intended null result. PC separates strongly from variance and from the specifically matched weak BRB moment M_3. It should add little to a strong BRB certificate that already makes the adverse shelf expensive. The benefit of PC is therefore not that it beats every BRB specification. The benefit is that it exposes the mechanism and remains interpretable when the operative BRB exponent is not obvious or is estimated unreliably.

The finite-sample comparison uses a one-dimensional nearest-neighbor estimator of loss-side varentropy. Let $Z_i = -X_i$ denote the observed loss magnitudes, restricted to observations with $X_i < 0$, and let n_- be the number of such loss observations. Let $R_{i,k}$ be the distance from Z_i to its kth nearest neighbor among the other loss observations, and define the local score proxy

$$L_{i,k} = \log\left(\frac{2(n_- - 1)R_{i,k}}{k}\right).$$

The raw variance of these local scores includes the usual nearest-neighbor log-spacing noise. Under the local Poisson approximation, this noise has variance $\psi_1(k)$, where ψ_1 is the trigamma function. We therefore use

$$s_k^2 = \frac{1}{n_- - 1}\sum_i (L_{i,k} - \bar{L}_k)^2, \qquad \widehat{V}_{-,k}^{\mathrm{kNN}} = \left[s_k^2 - \psi_1(k)\right]_+.$$

The table reports $k = 1, 2, 3$.

Table 7. Finite-sample behavior at window length $n = 1260$. The V_- rows are Monte Carlo summaries of oracle-score and bias-corrected kNN loss-side varentropy estimates. The raw M_λ rows use the analytic sampling RSE of the sample mean of $e^{-\lambda X}$. Among $k = 1, 2, 3$, $k = 3$ has the lowest RMSE for Law B and the smallest spurious estimate for Law A.

Law	Quantity	Population target	Mean estimate	SD estimate	RMSE
A	$\widehat{V}_-$, kNN $k = 1$	0.00	0.118	0.193	0.226
A	$\widehat{V}_-$, kNN $k = 2$	0.00	0.046	0.076	0.088
A	$\widehat{V}_-$, kNN $k = 3$	0.00	0.031	0.050	0.059
B	V_-, oracle-score sample	1.58	1.573	0.231	0.231
B	$\widehat{V}_-$, kNN $k = 1$	1.58	1.578	0.280	0.280
B	$\widehat{V}_-$, kNN $k = 2$	1.58	1.587	0.182	0.182
B	$\widehat{V}_-$, kNN $k = 3$	1.58	1.595	0.162	0.163
B	raw M_{60} sample mean	1.753	1.753	0.132	0.132
B	raw M_{80} sample mean	3.476	3.476	0.418	0.418
B	raw M_{100} sample mean	8.752	8.752	1.317	1.317

Table 7 fills the finite-sample gap in the BRB comparison at a long rolling-window length. It does not claim that every nonparametric estimate of V_- is more accurate than every estimate of M_λ. The point is narrower. In the matched shelf experiment, the nearest-neighbor score estimator with $k = 3$ is centered near the true adverse-law value $V_- = 1.58$, gives the smallest RMSE among $k = 1, 2, 3$, and assigns only a small residual value to Law A, whose loss side has zero population score dispersion. The raw negative-power moment, by contrast, becomes increasingly near-maximum as λ grows.

At $n = 1260$ both the kNN varentropy estimator and the raw BRB moment estimates are already fairly stable. Financial applications often use shorter effective samples, so the more relevant stress test is $n \in \{200, 300, 400\}$. In Law B the rare adverse shelf has probability three percent, so these windows contain only about six, nine, and twelve adverse-shelf observations on average. The smallest window is genuinely sparse: it has probability about 14.7% of observing three or fewer rare-shelf points.

Table 8. Small-sample rarity of the adverse shelf in Law B. Counts use the population shelf probability $p = 3\%$. The final column reports the expected number of loss-side observations, since $\mathbb{P}_B(X < 0) = 48\%$.

n	$\mathbb{E}N_{\mathrm{rare}}$	$\mathbb{P}(N_{\mathrm{rare}} = 0)$	$\mathbb{P}(N_{\mathrm{rare}} \le 3)$	$\mathbb{E}n_-$
200	6.0	0.23%	14.7%	96.0
300	9.0	0.01%	2.0%	144.0
400	12.0	0.00%	0.2%	192.0

Table 9. Small-sample kNN downside-varentropy estimation. For Law A the target is $V_- = 0$; for Law B the target is $V_- = 1.58$. The table shows that $k = 3$ is the best of $k = 1, 2, 3$ in these experiments: it has the lowest RMSE under both the null body-dominated law A and the adverse-hit law B.

n	k	A mean	A RMSE	B mean	B SD	B RMSE
200	1	0.267	0.603	1.614	0.707	0.708
200	2	0.110	0.238	1.687	0.444	0.457
200	3	0.073	0.158	1.788	0.388	0.440
300	1	0.225	0.481	1.604	0.577	0.578
300	2	0.089	0.189	1.641	0.379	0.384
300	3	0.058	0.122	1.673	0.332	0.345
400	1	0.203	0.429	1.594	0.499	0.499
400	2	0.084	0.175	1.619	0.330	0.332
400	3	0.054	0.113	1.639	0.292	0.298

The relevant comparison is not only the coefficient of variation of one statistic under one law. A risk screen must separate the adverse-hit law from the body-dominated law. For a statistic Q, define $\widehat{\Delta}_Q = \widehat{Q}_B - \widehat{Q}_A$ from independent samples under the two laws, and summarize its standardized separation $\mathbb{E}\widehat{\Delta}_Q / \operatorname{sd}(\widehat{\Delta}_Q)$. Table 10 shows that $\widehat{V}_-^{\mathrm{kNN}}$ with $k = 3$ gives the cleanest small-sample discrimination. At $n = 200$, its standardized separation is 4.11, compared with 2.09, 2.24, and 2.28 for raw M_{60}, M_{80}, and M_{100}. The weak matched moment M_3 is intentionally indifferent, and M_{20} has only modest separation.

Table 10. Small-sample discrimination between the adverse-hit law B and the body-dominated law A. $\mathbb{P}(B > A)$ is the Monte Carlo probability that the sample statistic is larger under B than under A. "Std. sep." is $\mathbb{E}(\widehat{Q}_B - \widehat{Q}_A) / \operatorname{sd}(\widehat{Q}_B - \widehat{Q}_A)$.

n	Statistic	$\mathbb{P}(B > A)$	Mean diff.	Std. sep.
200	$\widehat{V}_-^{\mathrm{kNN}}$, $k = 3$	99.8%	1.721	4.11
200	M_3	50.0%	0.000	0.03
200	M_{20}	80.2%	0.026	0.86
200	M_{60}	99.3%	0.696	2.09
200	M_{80}	99.7%	2.344	2.24
200	M_{100}	99.8%	7.508	2.28
300	$\widehat{V}_-^{\mathrm{kNN}}$, $k = 3$	100.0%	1.611	4.66
300	M_3	50.6%	0.000	0.04
300	M_{20}	85.5%	0.025	1.04
300	M_{60}	99.9%	0.694	2.56
300	M_{80}	100.0%	2.341	2.74
300	M_{100}	100.0%	7.498	2.79
400	$\widehat{V}_-^{\mathrm{kNN}}$, $k = 3$	100.0%	1.592	5.15
400	M_3	51.6%	0.000	0.07
400	M_{20}	89.6%	0.026	1.24
400	M_{60}	100.0%	0.700	2.96
400	M_{80}	100.0%	2.358	3.16
400	M_{100}	100.0%	7.548	3.21

Finally, Table 11 repeats the non-tail-model BRB regularization check at the most stressful window length, $n = 200$, for $\lambda = 80$. Raw M_{80} detects the adverse shelf, but its coefficient of variation is about 30%. Mild 99% winsorization preserves much of the signal but does not stabilize the statistic. Stronger 95% trimming stabilizes the estimate but moves it from the population value 3.476 to 0.979, which hides the adverse shelf. This is the practical estimation tradeoff: without a model for the tail, variance-reduction of a large-λ BRB statistic can remove the observations that carry the adverse-cell signal.

Table 11. Small-sample BRB regularization at $n = 200$ for Law B and $\lambda = 80$. The population value is $M_{80}(B) = 3.476$. Non-tail-model regularization either fails to stabilize the estimate or stabilizes it by removing the adverse-shelf signal.

Estimator	Mean estimate	SD estimate	CV
Raw sample mean	3.472	1.047	30.2%
99% winsorized	3.240	1.144	35.3%
95% winsorized	1.124	0.672	59.8%
99% trimmed	2.442	0.951	38.9%
95% trimmed	0.979	0.129	13.2%

These small-sample panels sharpen the PC/BRB distinction. They do not say that $\widehat{V}_-$ has smaller variance than every BRB moment. For example, M_{60} can have a slightly smaller coefficient of variation than $\widehat{V}_-^{\mathrm{kNN}}$. They show something more relevant for the mandate: the integrated loss-side score dispersion gives a more stable separation of the adverse-hit law from the body-dominated law in short samples, while weak BRB moments miss the difference and very aggressive BRB moments detect it through a noisy near-maximum functional.

Table 12. Regularized BRB estimators for Law B at $n = 1260$. Mild 99% winsorization preserves much of the signal but does little to reduce sampling variability. Stronger non-tail-model regularization stabilizes the statistic but moves the estimate back toward one, hiding the adverse shelf. Entries in parentheses are Monte Carlo coefficients of variation.

λ	Population $M_\lambda(B)$	Raw mean	Winsor 99%	Winsor 95%	Trim 99%	Trim 95%
60	1.753	1.752 (7.4%)	1.710 (7.9%)	0.961 (1.0%)	1.414 (8.6%)	0.942 (1.1%)
80	3.476	3.477 (11.8%)	3.298 (13.3%)	0.981 (1.2%)	2.356 (16.1%)	0.958 (1.3%)
100	8.752	8.758 (15.2%)	8.043 (17.9%)	1.010 (1.4%)	5.082 (23.2%)	0.981 (1.5%)

Table 12 is not meant to defeat every possible BRB estimator. It brackets a structural tradeoff. Without an explicit tail model, regularization that reduces the variance of a large-λ negative-power moment can also remove the observations that contain the adverse-shelf signal. A model-based tail repair may preserve the signal, but then the estimator has reintroduced a tail model or tail-index assumption. PC avoids that particular choice by using V_- as an integrated witness and by reporting the body/hit composition directly.

The controlled experiment therefore supports the paper's intended thesis. PC and aggressive BRB target overlapping adverse mass. When BRB is correctly specified, strong, and well estimated, PC should mostly agree with it. PC adds value in the regimes where BRB is weak, misspecified, or statistically unstable, and it adds value even alongside BRB by resolving scalar left-tail risk into body accumulation and adverse-hit exposure.

V Classical score-tail calibrations

This appendix gives the calculations behind Table 2. In all symmetric examples, the sign of X is independent of $|X|$, and $\iota(X)$ is an even function. Hence

$$p_- = \frac{1}{2}, \qquad h_- = h, \qquad V_- = \mathrm{VE}(X).$$

For $a = k\sqrt{V_-}$, the distribution-free benchmark is therefore

$$\frac{1}{2(1+k^2)}.$$

V-A Gaussian

If $X \sim N(0, \sigma^2)$ and $Z = X/\sigma$, then

$$\iota(X) = \frac{1}{2}\log(2\pi\sigma^2) + \frac{Z^2}{2}, \qquad h = \frac{1}{2}\log(2\pi e\sigma^2), \qquad \mathrm{VE}(X) = \frac{1}{2}.$$

Thus

$$\iota(X) - h = \frac{1}{2}(Z^2 - 1).$$

The loss-side high-information event at $a = k\sqrt{1/2}$ is

$$\left\{ Z \le -\sqrt{1+\sqrt{2}\,k} \right\},$$

and the exact probability is

$$\mathbb{P}(A_k^-) = \Phi\left(-\sqrt{1+\sqrt{2}\,k}\right). \tag{93}$$

V-B Laplace

For the symmetric Laplace density $f(x) = (2b)^{-1}e^{-|x|/b}$,

$$\iota(X) = \log(2b) + \frac{|X|}{b}, \qquad \frac{|X|}{b} \sim \mathrm{Exp}(1), \qquad \mathrm{VE}(X) = 1.$$

Therefore

$$\iota(X) - h = \frac{|X|}{b} - 1,$$

and

$$\mathbb{P}(A_k^-) = \frac{1}{2}e^{-(1+k)}. \tag{94}$$

V-C Student-t

Let $Z \sim t_\nu$ with density

$$f_\nu(z) = c_\nu \left(1 + \frac{z^2}{\nu}\right)^{-(\nu+1)/2}, \qquad c_\nu = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}.$$

The scale parameter is irrelevant for varentropy. The self-information is

$$\iota(Z) = -\log c_\nu + \frac{\nu+1}{2}\log\left(1 + \frac{Z^2}{\nu}\right).$$

Using

$$\frac{Z^2}{\nu + Z^2} \sim \mathrm{Beta}\left(\frac{1}{2}, \frac{\nu}{2}\right),$$

one obtains

$$\mu_\nu := \mathbb{E}\log\left(1 + \frac{Z^2}{\nu}\right) = \psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)$$

and

$$\mathrm{VE}(t_\nu) = \frac{(\nu+1)^2}{4}\left[\psi_1\left(\frac{\nu}{2}\right) - \psi_1\left(\frac{\nu+1}{2}\right)\right], \tag{95}$$

where ψ and ψ_1 are the digamma and trigamma functions. The event $\iota(Z) - h \ge k\sqrt{\mathrm{VE}(t_\nu)}$ is equivalent to $|Z| \ge r_{\nu,k}$, where

$$r_{\nu,k} = \left[\nu\left(\exp\left(\mu_\nu + \frac{2k\sqrt{\mathrm{VE}(t_\nu)}}{\nu+1}\right) - 1\right)\right]^{1/2}. \tag{96}$$

Thus

$$\mathbb{P}(A_k^-) = \mathbb{P}(t_\nu \le -r_{\nu,k}). \tag{97}$$

The table evaluates this expression for $\nu = 1, 3, 5, 10$.

V-D Normal inverse Gaussian

For the normal inverse Gaussian law we use the parameterization of [13]

$$f(x) = \frac{\alpha\delta}{\pi\sqrt{\delta^2 + (x-\mu)^2}} K_1\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)\exp\left\{\delta\gamma + \beta(x-\mu)\right\}, \qquad \gamma = \sqrt{\alpha^2 - \beta^2},$$

where K_1 is the modified Bessel function of the second kind. The row in Table 2 uses the symmetric case $\alpha = \delta = 1$, $\beta = \mu = 0$. The entropy and varentropy are computed by one-dimensional quadrature,

$$h = \int_{-\infty}^{\infty} -\log f(x) f(x)\,dx, \qquad \mathrm{VE} = \int_{-\infty}^{\infty} \left(-\log f(x) - h\right)^2 f(x)\,dx,$$

which gives approximately

$$h \approx 1.36976, \qquad \mathrm{VE} \approx 0.89749.$$

For each k, the score threshold is the positive solution r_k of

$$-\log f(r_k) = h + k\sqrt{\mathrm{VE}},$$

and the exact loss-side probability is

$$\mathbb{P}(A_k^-) = \int_{r_k}^{\infty} f(x)\,dx, \tag{98}$$

because of symmetry. Numerically this gives the NIG row of Table 2. The point of this example is not a closed-form identity but a finance-standard semi-heavy density with finite variance and explicit log-scores.

V-E Closed-form loss-side varentropy for exponential-power / generalized Gaussian.

Recall the digamma and trigamma functions $\psi = \Gamma'/\Gamma$, $\psi_1 = \psi'$, and the regularized upper incomplete gamma function $Q(s,z) = \Gamma(s,z)/\Gamma(s)$.

Theorem V.1. *Let X be symmetric about 0 with a density in one of the following scale families, so that $p_- = \frac{1}{2}$, $h_- = h$, and $V_- = \mathrm{VE}(X)$. If*

$$f(x) = \frac{\beta}{2s\,\Gamma(1/\beta)}\,\exp\!\big(-|x/s|^{\beta}\big), \qquad \beta > 0,\ s > 0,$$

then, writing $W := |X/s|^{\beta} \sim \mathrm{Gamma}(1/\beta, 1)$ and $c_\beta := \log\!\big(2s\,\Gamma(1/\beta)/\beta\big)$,

$$\imath(X) = c_\beta + W, \qquad h = c_\beta + \tfrac{1}{\beta}, \qquad \mathrm{VE}(X) = \mathrm{Var}(W) = \frac{1}{\beta}. \tag{99}$$

Moreover the excess-information law is exact:

$$\mathbb{P}\{\imath(X) - h \ge a \mid X < 0\} \;=\; Q\!\Big(\tfrac{1}{\beta},\ \tfrac{1}{\beta} + a\Big), \qquad a \ge -\tfrac{1}{\beta}, \tag{100}$$

an incomplete-gamma function with no asymptotic-tail approximation.

The distribution-free hit budget of Theorem III.15 is therefore available in closed form through V_-, and in the exponential-power family the exact hit probability (100) *can be compared term by term against the Cantelli bound $V_-/(V_- + a^2) = (1/\beta)/(1/\beta + a^2)$.*

Proof. By definition $\imath(x) = -\log f(x) = c_\beta + |x/s|^{\beta}$, which is the first identity in (99) with $W = |X/s|^{\beta}$. The folded variable $|X|/s$ has density $\big(\beta/\Gamma(1/\beta)\big)e^{-u^{\beta}}$ on $u > 0$. The change of variables $W = u^{\beta}$, $u = W^{1/\beta}$, $\mathrm{d}u = \frac{1}{\beta}W^{1/\beta-1}\,\mathrm{d}W$, gives the density of W as $\Gamma(1/\beta)^{-1}W^{1/\beta-1}e^{-W}$, i.e. $W \sim \mathrm{Gamma}(1/\beta, 1)$. Hence $\mathbb{E}W = \mathrm{Var}\,W = 1/\beta$, which yields h and $\mathrm{VE}(X) = \mathrm{Var}(\imath(X)) = \mathrm{Var}(W) = 1/\beta$. Because the additive constant c_β cancels in centering, the excess information is $\imath(X) - h = W - \frac{1}{\beta}$, and

$$\mathbb{P}\{\imath(X) - h \ge a \mid X < 0\} = \mathbb{P}\big(W \ge \tfrac{1}{\beta} + a\big) = Q\big(\tfrac{1}{\beta}, \tfrac{1}{\beta} + a\big),$$

the symmetry of f making the loss-side conditional law of W identical to its unconditional law. This is (100). (The cases $\beta = 2$ and $\beta = 1$ recover the Gaussian value $\mathrm{VE} = \frac{1}{2}$ and the Laplace value $\mathrm{VE} = 1$ of Section III-J.) □

VI Tail-regular calibrations

This appendix records the few tail-regular facts that remain useful after the main theorem has been rewritten in finite-threshold form. They are calibrations: they show how the adverse-set certificate collapses to familiar heavy-tail and single-big-jump formulas in a regular-varying model.

VI-A Regular variation gives an exponential score tail

Assume the left density tail satisfies

$$f(-x) \sim A_{-} x^{-(\beta+1)}, \qquad x \to \infty, \tag{101}$$

with survival-tail exponent $\beta > 0$. Then

$$\bar{F}_{-}(x) \sim \frac{A_{-}}{\beta} x^{-\beta}, \qquad \imath(-x) = (\beta+1)\log x + O(1).$$

Consequently the loss-side score tail has exponential rate

$$\mathbb{P}\{X < 0, \imath(X) \geq t\} = \exp\{-\lambda t + O(1)\}, \qquad \lambda = \frac{\beta}{\beta+1}. \tag{102}$$

Thus a regularly varying loss tail is one sufficient mechanism for a Pareto-score envelope with $\lambda = \beta/(\beta+1)$. The main text uses the reverse language carefully: a Pareto-score envelope is a score-tail assumption, not a literal regular-variation assumption on returns.

VI-B Tail-excess self-information

Under (101), the excess score beyond a large loss threshold has an exponential limit:

$$\imath(X) - \imath(-c) \,\Big|\, \{X \leq -c\} \Rightarrow \mathrm{Exp}\left(\frac{\beta}{\beta+1}\right), \qquad c \to \infty. \tag{103}$$

In particular,

$$\mathrm{Var}\big(\imath(X) - \imath(-c) \mid X \leq -c\big) \longrightarrow \left(\frac{\beta+1}{\beta}\right)^2. \tag{104}$$

This is the sharp special-case statement behind the paper's interpretation of varentropy-like quantities as score-tail dispersion measures.

VI-C Single-big-jump recovery

If the left tail is subexponential, then for fixed finite T,

$$\mathbb{P}(\underline{S}_T \leq -c) \sim T\bar{F}_{-}(c), \qquad c \to \infty. \tag{105}$$

This is the classical one-large-adverse-increment limit [14]. It is now just a calibration of the set-indexed theorem with $A = (-\infty, -c]$: the adverse-hit term becomes $T\bar{F}_{-}(c)$ and the body term is negligible after excluding adverse increments. In continuous-time Lévy models the analogous calibration replaces the one-period probability by the large-negative-jump intensity; if $\nu(-\infty, -c) \sim Cc^{-\beta}$, then under standard subexponential one-large-jump hypotheses

$$\mathbb{P}\left(\inf_{0\leq t\leq T} X_t \leq -c\right) \sim TCc^{-\beta}.$$

Symmetric α_{st}-stable processes are the self-similar fixed points of this picture [15], with $\beta = \alpha_{\mathrm{st}}$ and $\inf_{t\leq T} X_t \overset{d}{=} T^{1/\alpha_{\mathrm{st}}} \inf_{t\leq 1} X_t$.

References

[1] A. V. Stankyavichyus, "Varentropy: Overview, computational routes, and structural decomposition," *preprint*, 2026. [Online]. Available: https://anatolyvitold.com/preprints/varentropy_decomposition.pdf

[2] A. V. Stankyavichyus, "Varentropy of stable laws: A constructive formula at rational index," *preprint*, 2026. [Online]. Available: https://anatolyvitold.com/preprints/varentropy_stable_laws.pdf

[3] E. Busseti, E. K. Ryu, and S. P. Boyd, "Risk-constrained kelly gambling," *The Journal of Investing*, vol. 25, pp. 118–134, 2016. [Online]. Available: https://doi.org/10.3905/joi.2016.25.3.118

[4] P. Billingsley, *Convergence of probability measures* (Wiley Series in Probability and Statistics), 2nd ed. Nashville, TN: John Wiley & Sons, 1999. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316962

[5] O. Vasicek, “A test for normality based on sample entropy,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 38, no. 1, pp. 54–59, 1976. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1976.tb01566.x

[6] L. Kozachenko and N. Leonenko, “Sample estimate of the entropy of a random vector,” *Problems of Information Transmission*, vol. 23, pp. 9–16, 2 1987.

[7] I. R. Savage, “Probability inequalities of the Tchebycheff type,” *Journal of Research of the National Bureau of Standards-B. Mathematics and Mathematical Physics*, vol. 65B, no. 3, pp. 211–222, 1961. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/jres/65b/jresv65bn3p211_a1b.pdf

[8] L. L. Cam, “An approximation theorem for the Poisson binomial distribution.,” *Pacific Journal of Mathematics*, vol. 10, no. 4, pp. 1181–1197, 1960.

[9] A. D. Barbour and P. Hall, “On the rate of poisson convergence,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 95, no. 3, pp. 473–480, 1984.

[10] A. DasGupta, *Asymptotic theory of statistics and probability* (Springer Texts in Statistics), 2008th ed. New York, NY: Springer, 2008.

[11] I. Shevtsova, “On the absolute constants in the berry–esseen type inequalities for identically distributed summands,” 2011. arXiv: 1111.6554. [Online]. Available: https://doi.org/10.48550/arXiv.1111.6554

[12] A. B. Tsybakov, *Introduction to Nonparametric Estimation* (Springer Series in Statistics). New York, NY: Springer, 2009. [Online]. Available: https://link.springer.com/book/10.1007/978-3-030-52915-4

[13] O. E. Barndorff-Nielsen, “Normal inverse gaussian distributions and stochastic volatility modelling,” *Scandinavian Journal of Statistics*, vol. 24, no. 1, pp. 1–13, 1997. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9469.00045

[14] S. Foss, D. Korshunov, and S. Zachary, *An introduction to heavy-tailed and subexponential distributions* (Springer Series in Operations Research and Financial Engineering), 2011th ed. New York, NY: Springer, 2011. [Online]. Available: https://doi.org/10.1007/978-1-4419-9473-8

[15] J. P. Nolan, *Univariate stable distributions* (Springer Series in Operations Research and Financial Engineering), 2020th ed. Cham, Switzerland: Springer Nature, 2020. [Online]. Available: https://link.springer.com/book/10.1007/978-3-030-52915-4