

UC Berkeley

SEMM Reports Series

Title

Rare-Event Probability Scaling and Extrapolation with Renormalization Group Flow

Permalink

<https://escholarship.org/uc/item/3hh3x1np>

Author

Wang, Ziqi

Publication Date

2026-09-21

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Report No.
UCB/SEMM-[2026/01]

Structural Engineering
Mechanics and Materials

Rare-Event Probability Scaling and
Extrapolation with Renormalization Group Flow

By

Ziqi Wang

[September 2026]

Department of Civil and Environmental Engineering
University of California, Berkeley

Rare-Event Probability Scaling and Extrapolation with Renormalization Group Flow

Ziqi Wang^{a,*}

^a*Department of Civil and Environmental Engineering, University of California, Berkeley, United States*

Abstract

We formulate rare-event probability as a scaling limit of a cumulant generating function. The corresponding renormalization group flow enables extrapolation to the rare-event limit. For typical model choices, the flow admits a universal power-law scaling. We demonstrate this approach with an example. The results, however, lack significant practical utility. This work is reported for theoretical interest.

Keywords: rare-event simulation, renormalization group, uncertainty quantification

1. Problem Statement

Rare-event probability estimation [1–3] has important applications across science and engineering. Examples include quantifying the risk of natural disasters, financial crises, network failures, and social instabilities. Let a Bernoulli random variable $B \in \{0, 1\}$ represent the rare event of interest, with $p \equiv \mathbb{P}(B = 1)$ denoting the probability to be estimated. We use a limit-state function to represent the rare event:

$$B = \mathbb{1}_{\leq 0}(G(\mathbf{X})),$$

where $\mathbb{1}_{\leq 0}(g)$ is the indicator function for $g \leq 0$, $G : \mathbb{R}^n \rightarrow \mathbb{R}$ is the limit-state function modeling the underlying process that triggers the rare event, and $\mathbf{X} \in \mathbb{R}^n$ is a random vector with known distribution $\mathbb{P}_{\mathbf{X}}$. In contrast with the uninformative Bernoulli variable B , the fine details of G and $\mathbb{P}_{\mathbf{X}}$ make efficient probabilistic computations possible. Analytical solutions for the probability, alternatively represented as $\mathbb{E}[\mathbb{1}_{\leq 0}(G(\mathbf{X}))]$, may be available for simple G and $\mathbb{P}_{\mathbf{X}}$. For nontrivial problems where G is nonlinear and $\mathbb{P}_{\mathbf{X}}$ is high-dimensional, the most widely used approach is importance sampling coupled with surrogate modeling in various forms [4–7]. This work introduces an alternative angle to this problem through a statistical physics interpretation of rare-event probability estimation. The key idea is to cast probability estimation as a crossover scaling of a cumulant generating function. The scale transformation admits two trivial fixed points, one corresponding to the rare-event limit. When the cumulant generating function is close to this fixed point, the universal scaling law enables extrapolation of the rare-event probability.

2. Renormalization Group Ansatz for Rare-event Probability Estimation

We first replace the indicator function $\mathbb{1}_{\leq 0}(g)$ by a complementary cumulative distribution function $\bar{F}(sg)$ with a control parameter $s > 0$, where $\bar{F}$ can from any probability distribution supported on $\mathbb{R}$ and symmetric around 0. As $s \rightarrow \infty$, the fine details of g are suppressed and only the sign of g is preserved in $\bar{F}(sg)$, recovering the indicator function for the rare event. In contrast, as $s \rightarrow 0^+$, the fine details of g are again washed out, with $\bar{F}(sg) \rightarrow 0.5$, a trivial constant.

Next, we examine the cumulant generating function of $\bar{F}(sG(\mathbf{X}))$, assuming it exists:

$$K(s, t) := \log \mathbb{E} \left[e^{t \bar{F}(sG(\mathbf{X}))} \right], \quad t \in \mathbb{R}.$$

*Corresponding author

Email address: ziquwang@berkeley.edu (Ziqi Wang)

The function $K(s, t)$ fully characterizes the statistics of $\bar{F}(sG(\mathbf{X}))$ and is therefore sufficient for estimating the rare-event probability. In the limit $s \rightarrow \infty$, the fixed point of $K(s, t)$ is the Bernoulli cumulant generating function,

$$K_\infty(t) := \lim_{s \rightarrow \infty} K(s, t) = \log(1 - p + pe^t),$$

where p is the rare-event probability of interest. In the opposite limit, $s \rightarrow 0^+$, the fixed point is the uninformative form

$$K_0(t) := \lim_{s \rightarrow 0^+} K(s, t) = \frac{t}{2}.$$

Thus, as s varies from 0 to large values, $K(s, t)$ flows from $K_0(t)$ to $K_\infty(t)$. We next investigate the structure of this flow. Under mild conditions, we obtain the scaling laws (see [Appendix A](#)):

$$\begin{aligned} \text{small } s: \quad & K(s, t) - K_0(t) \sim s, & \text{if } \mathbb{E}[G(\mathbf{X})] \text{ exists and is nonzero,} \\ \text{large } s: \quad & K(s, t) - K_\infty(t) \sim s^{-2}, & \text{if } G(\mathbf{X}) \text{ has a density } f \text{ around 0 and } f'(0) \neq 0. \end{aligned} \quad (1)$$

Other cases are discussed in [Appendix A](#). The renormalization group (RG) ansatz for approximating $K(s, t)$ attracted to the rare-event fixed point is

$$K(s, t) = K_\infty(t) + (K(s_c, t) - K_\infty(t)) \frac{s_c^2}{s^2}, \quad (2)$$

where s_c is a characteristic scale beyond which the law s^{-2} dominates the attraction toward $K_\infty(t)$. To construct a numerical extrapolation for estimating the probability p , the only parameter to be specified is the characteristic scale s_c . The linear scaling law in the small- s regime can serve as a diagnostic for selecting s_c .

3. Computational Approach

Building on Section 2, one can design various computational approaches for extrapolating rare-event probabilities. Here we introduce a simple scheme to prove the concept, without attempting to optimize implementation details.

Assume we have numerical estimates of $K(s, t)$ over a grid of $(s \neq 0, t \neq 0)$ values. These estimates can be obtained using samples from $\mathbb{P}_{\mathbf{X}}$, importance sampling, control variates, or other variance reduction techniques. The main computational cost typically lies in evaluating G for each sample. The grid resolution in (s, t) is usually not a bottleneck, since $K(s, t)$ is obtained by postprocessing existing samples.

Next, we identify the characteristic scale s_c at which the scaling law toward $K_\infty(t)$ becomes effective. Since the $K(s, t)$ estimates for small s are more reliable, we consider a complementary problem: the scale at which the linear scaling law toward the trivial fixed point $K_0(t)$ fails substantially. The linear law suggests

$$\frac{K(s^-, t) - t/2}{K(s, t) - t/2} = \frac{s^-}{s},$$

where $s^- > 0$ is the smallest scale at which $K(s, t)$ is estimated. This can be reformulated into a diagnostic metric for the breakdown of the linear law, expressed as

$$\delta(s; s^-, t) = \left| \log \left(\frac{s(K(s^-, t) - t/2)}{s^-(K(s, t) - t/2)} \right) \right|. \quad (3)$$

This metric as a function of s would reveal the scale at which the linear law fails substantially. As a simple rule of thumb, one could use a threshold of $\log(1 + 50\%) \approx 0.4$. The metric is generally insensitive to t ; a conservative approach is to take the maximum value over a range of t .

Let $K_{\text{data}}(s, t)$ denote the numerical estimates of $K(s, t)$. From Eq. (2), for a given t ,

$$H(s, t, p) := s^2 (K_{\text{data}}(s, t) - \log(1 - p + pe^t)) \approx \text{constant}, \quad s \geq s_c.$$

An estimate p' should make $H(s, t, p')$ as constant as possible with respect to s for each t , that is, the variability of H across s should be small. This leads to the optimization formulation

$$p_{\text{RG}} = \arg \min_{p'} \sum_t \sum_s (H(s, t, p') - \bar{H}(t, p'))^2, \quad (4)$$

where $\bar{H}(t, p')$ is the average of $H(s, t, p')$ over s . In practice, a weight $1/\text{Var}[K_{\text{data}}(s, t)]$ can be applied to downweight unreliable estimates.

4. A Proof-of-Concept Example

Consider a Duffing oscillator driven by band-limited Gaussian white noise,

$$\ddot{z}(\tau) + 2\xi\omega_n\dot{z}(\tau) + \omega_n^2 z(\tau) + \beta z^3(\tau) = a(\tau),$$

where $\xi = 0.05$, $\omega_n = 10 \text{ rad s}^{-1}$, and $\beta = 2000 \text{ m}^{-2}\text{s}^{-2}$. The excitation $a(\tau)$ is Gaussian white noise with a cutoff angular frequency $30\pi \text{ rad s}^{-1}$ and spectral intensity $1/(30\pi) \text{ m}^2 \text{ s}^{-3} \text{ rad}^{-1}$, chosen so that the marginal variance of $a(\tau)$ is unity. We study a first-passage event defined by the limit-state function

$$G(\mathbf{x}) = z_0 - \sup_{\tau \in [0, 10]} |z(\tau; \mathbf{x})|,$$

where $\mathbf{x}$ denotes a finite-dimensional representation of $a(\tau)$, and $z_0 > 0$ is a threshold.

First, we numerically verify the scaling laws in Eq. (1) using $z_0 = 0.07$ and Monte Carlo simulation with 10^5 samples. Figure 1 compares the Monte Carlo estimates of $K(s, t)$ in the small- and large- s regimes with the analytical scaling predictions. Next, we evaluate the diagnostic metric associated with the linear law toward $K_0(t)$, as shown in Figure 2. Finally, to demonstrate the RG-based extrapolation, we perform a rolling prediction using $K_{\text{data}}(s, t)$ over $[s_c, s]$ for increasing s and $t \in [-1, 1]$. Figure 3 presents the results alongside the sequence of smoothed probability estimates $\hat{p}(s)$.

5. Conclusion

We developed a scaling theory for rare-event probability extrapolation. Its practical utility is limited by a fundamental trade-off between the increasing accuracy of the theory and the decreasing accuracy of the data in the regime of interest.

References

- [1] G. Rubino, B. Tuffin, Rare event simulation using Monte Carlo methods, John Wiley & Sons, 2009.
- [2] J. A. Bucklew, J. Bucklew, Introduction to rare event simulation, Vol. 5, Springer, 2004.
- [3] R. Y. Rubinstein, D. P. Kroese, Simulation and the Monte Carlo method, John Wiley & Sons, 2008.
- [4] A. Der Kiureghian, Structural and System Reliability, Cambridge University Press, 2022.
- [5] B. Echard, N. Gayton, M. Lemaire, AK-MCS: An active learning reliability method combining Kriging and Monte Carlo simulation, Structural Safety 33 (2) (2011) 145–154.
- [6] B. Sudret, Meta-models for structural reliability and uncertainty quantification, arXiv preprint arXiv:1203.2062 (2012).
- [7] J. Xian, Z. Wang, A physics and data co-driven surrogate modeling method for high-dimensional rare event simulation, Journal of Computational Physics 510 (2024) 113069.

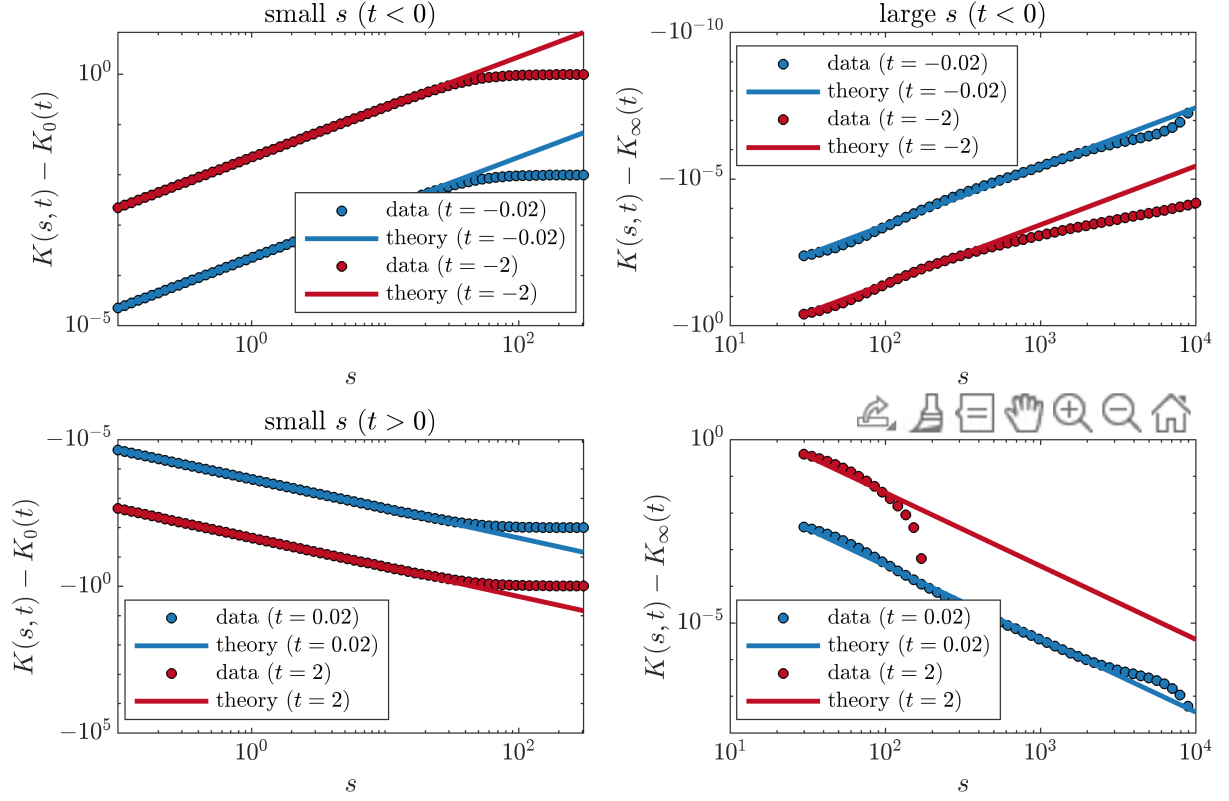


Figure 1: **Verification of the scaling laws.** The linear scaling law toward the trivial fixed point $K_0(t)$ is robust and largely insensitive to t . The s^{-2} scaling toward the rare-event fixed point $K_\infty(t)$ is more accurate for small $|t|$. The derivation in [Appendix A](#) assumes $|t|$ to be relatively small. The precise range of *small* $|t|$ depends on the rate at which $\mathbb{E}[\bar{F}(sG(\mathbf{X}))]$ approaches $1/2$ or p at the respective scaling limits. In general, faster convergence of $\mathbb{E}[\bar{F}(sG(\mathbf{X}))]$ allows a wider feasible range of $|t|$.

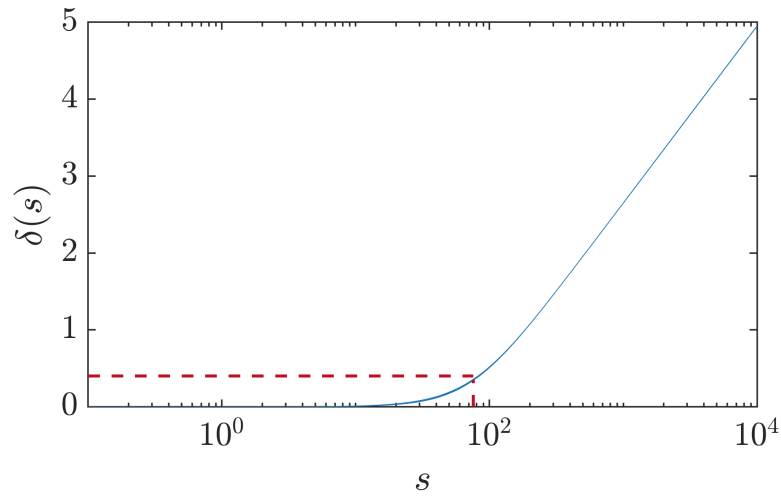


Figure 2: **Diagnostic metric for the breakdown of the linear scaling law.** The figure overlays $\delta(s)$ curves for of a mesh of 100 uniformly spaced $t_i \in [-2, 2]$. It is seen that $\delta(s)$ is not sensitive to t . As expected, the linear scaling law breaks down as the scale increases. The dashed line marks $s_c \approx 75$ identified by the rule of thumb $\Delta(s_c) > 0.4$.

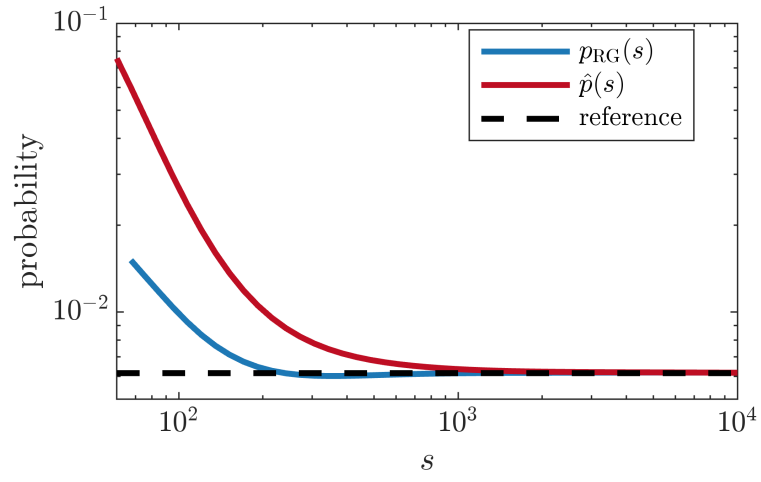


Figure 3: **Rolling RG extrapolation.** The RG-based prediction precedes the smoothed probability estimate $\hat{p}(s)$. However, long-range extrapolation seems to be challenging due to the trade-off between the increasing precision of the RG theory and the decreasing precision in the data at large s .

Appendix A. Scaling Laws

Rare-event limit

We denote the pushforward measure of $\mathbb{P}_{\mathbf{X}}$ under the limit-state function $G : \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$\mathbb{P}_G(B) := \mathbb{P}_{\mathbf{X}}(G^{-1}(B)), \quad B \in \mathcal{B},$$

where $G^{-1}(B) := \{\mathbf{x} \in \mathbb{R}^n : G(\mathbf{x}) \in B\}$ is the preimage of B , and $\mathcal{B}$ is the Borel sigma-algebra on $\mathbb{R}$. Consider the dilation map $T_s : \mathbb{R} \rightarrow \mathbb{R}$, $T_s(g) = sg$, and the corresponding dilated measure

$$\mu_s(B) := \mathbb{P}_G(T_s^{-1}(B)) = \mathbb{P}_G(\{g \in \mathbb{R} : sg \in B\}), \quad B \in \mathcal{B}.$$

As $s \rightarrow \infty$, $T_s^{-1}(B)$ shrinks toward $\{0\}$, suggesting that the scaling behavior toward the rare-event limit is determined by the property of $\mathbb{P}_G$ around 0, provided that such a neighborhood carries nonzero measure.

Let $\bar{F}(sg)$ represent a smoothed indicator function, where $\bar{F}$ is a generic complementary cumulative distribution function corresponding to a symmetric probability distribution centered at zero. The error in the probability estimate is

$$\Delta_p(s) := \int_{\mathbb{R}} (\bar{F}(sg) - \mathbb{1}_{\leq 0}(sg)) \mathbb{P}_G(dg) = \int_{\mathbb{R}} (\bar{F}(z) - \mathbb{1}_{\leq 0}(z)) \mu_s(dz).$$

Assume that $\mathbb{P}_G$ has a density f in a neighborhood of 0 that is sufficiently smooth (e.g., $f \in C^2$). Then the dilated measure μ_s has density

$$f_s(z) = \frac{1}{s} f\left(\frac{z}{s}\right).$$

Expanding $f(z/s)$ around 0 gives

$$f\left(\frac{z}{s}\right) = f(0) + \frac{z}{s} f'(0) + \frac{z^2}{2s^2} f''(0) + O(s^{-3}).$$

Substituting into the expression for $\Delta_p(s)$ yields

$$\begin{aligned} \Delta_p(s) &= \frac{1}{s} \left(f(0) \underbrace{\int_{\mathbb{R}} (\bar{F}(z) - \mathbb{1}_{\leq 0}(z)) dz}_{=: M_0} + \frac{f'(0)}{s} \underbrace{\int_{\mathbb{R}} z(\bar{F}(z) - \mathbb{1}_{\leq 0}(z)) dz}_{=: M_1} + \frac{f''(0)}{2s^2} M_2 + O(s^{-3}) \right) \\ &= \frac{f(0)M_0}{s} + \frac{f'(0)M_1}{s^2} + O(s^{-3}). \end{aligned}$$

Due to the symmetry $\bar{F}(z) + \bar{F}(-z) = 1$, we have $M_k = 0$ for $k = 0, 2, 4, \dots$. Thus,

$$\Delta_p(s) = \frac{f'(0)M_1}{s^2} + O(s^{-4}).$$

We express M_1 as

$$M_1 = \int_{-\infty}^0 z(\bar{F}(z) - 1) dz + \int_0^{\infty} z\bar{F}(z) dz = \int_0^{\infty} z(F(-z) + \bar{F}(z)) dz.$$

The symmetry $F(-z) = \bar{F}(z)$ further leads to

$$M_1 = 2 \int_0^{\infty} z \bar{F}(z) dz > 0,$$

so M_1 is strictly positive for any continuous nontrivial $\bar{F}(z)$. In fact, the integrand $z(F(-z) + \bar{F}(z))$ is nonnegative for all $z > 0$, so $M_1 \geq 0$ even without assuming symmetry, and strict positivity follows for nondegenerate $\bar{F}(z)$. For concrete examples of M_1 , expanding the sigmoid smoothing function as a geometric series yields

$$M_1 = 2 \int_0^{\infty} z \sum_{k=1}^{\infty} (-1)^{k+1} e^{-kz} dz = \sum_{k=1}^{\infty} \frac{2}{k^2} = \zeta(2) = \frac{\pi^2}{6}.$$

The standard Gaussian CDF yields

$$M_1 = 2 \int_0^\infty z \Phi(-z) dz = \int_0^\infty z^2 \phi(z) dz = \frac{1}{2}.$$

Finally, the scaling of $\Delta_p(s)$ is summarized as:

$$\Delta_p(s) \sim \frac{f^{(m)}(0) M_m}{m!} s^{-(m+1)}, \quad m := \min\{k \in \{1, 3, 5, \dots\} : f^{(k)}(0) \neq 0\}.$$

A corner case occurs when f is uniform around 0, leading to a decay faster than any power law, making extrapolation infeasible.

If $\mathbb{P}_G$ is atomic and supported on $\{g_i\}$, let

$$\mathbb{P}_G(B) = \sum_i a_i \mathbb{1}_{g_i \in B}, \quad a_i > 0, \quad \sum_i a_i = 1.$$

Then $\Delta_p(s)$ becomes

$$\begin{aligned} \Delta_p(s) &= -\frac{1}{2} \mathbb{P}_G(\{0\}) + \sum_{g_i < 0} (a_i \bar{F}(sg_i) - a_i) + \sum_{g_j > 0} a_j \bar{F}(sg_j) \\ &= -\frac{1}{2} \mathbb{P}_G(\{0\}) + \sum_{g_i \neq 0} \text{sign}(g_i) a_i \bar{F}(s|g_i|) \end{aligned}$$

This expression shows that an irreducible bias arises if $\mathbb{P}_G(\{0\}) > 0$, reflecting the smooth transition of the surrogate indicator function near the discontinuity of the genuine indicator function. It also implies that no universal scaling law exists for general atomic measures, since the decay of $\Delta_p(s)$ depends on the detailed atomic structure of $\mathbb{P}_G$. A special case occurs when the smallest nonzero atom,

$$g^- := \min_{g_i \neq 0} |g_i|,$$

is well separated from the others, leading to an approximate scaling law

$$\Delta_p(s) + \frac{1}{2} \mathbb{P}_G(\{0\}) \sim \bar{F}(sg^-).$$

This law may not be useful as it decays too rapidly for typical smoothed indicator functions.

We extend the scaling law of $\Delta_p(s)$ to the cumulant generating function. This transforms pointwise estimation of small probabilities to a spectral characterization of the underlying measure. The cumulant generating function serves as the analogue of free energy in the classical renormalization group framework. We consider the discrepancy between $K(s, t)$ and its Bernoulli limit $K_\infty(t)$:

$$\begin{aligned} \Delta_K(s, t) &:= K(s, t) - K_\infty(t) \\ &= \log \left(1 + \frac{\mathbb{E} \left[e^{t\bar{F}(G(\mathbf{X}); s)} \right] - (1-p) - pe^t}{1-p+pe^t} \right) \\ &= \log \left(1 + \frac{\sum_{k=1}^\infty \frac{t^k}{k!} (\mathbb{E} [\bar{F}^k(G(\mathbf{X}); s)] - p)}{1-p+pe^t} \right) \\ &= \log \left(1 + \frac{t \Delta_p(s) + \frac{t^2}{2} (\mathbb{E} [\bar{F}^2] - p) + O(t^3)}{1-p+pe^t} \right) \\ &\approx \frac{t}{1-p+pe^t} \Delta_p(s). \end{aligned}$$

Given that $|t|$ is relatively small, the linearization is expected to be accurate as $p \approx 0$ and $\Delta_p \approx 0$. Therefore, $\Delta_K(s, t)$ inherits the scaling of $\Delta_p(s)$:

$\Delta_K(s, t) \sim s^{-(m+1)}$, if $\mathbb{P}_G$ has a density f near 0 with the smallest odd $m \geq 1$ yielding $f^{(m)}(0) \neq 0$,

$\Delta_K(s, t) + \frac{1}{2} A(t) \mathbb{P}_G(\{0\}) \sim \bar{F}(sg^-)$, if $\mathbb{P}_G$ is atomic with a spectral gap g^- ,

where $A(t)$ is $t/(1-p+pe^t)$.

Uninformative limit

We examine the behavior of the smoothed probability as $s \rightarrow 0^+$. Redefine

$$\Delta_p(s) := \hat{p}(s) - \hat{p}(0) = \int_{\mathbb{R}} \bar{F}(sg) \mathbb{P}_G(dg) - \frac{1}{2}.$$

Treating sg as a single variable, the small- sg expansion of $\bar{F}(sg)$ is

$$\bar{F}(sg) = \bar{F}(0) - sg \pi(0) - \frac{s^2 g^2}{2} \pi''(0) + O(s^3 g^3),$$

where $\pi(\cdot)$ is the probability density function associated with $\bar{F}$. For typical choices of $\bar{F}$, such as Gaussian or sigmoid, we have $\pi(0) > 0$ and $\pi''(0) = 0$. It follows that

$$\begin{aligned} \Delta_p(s) &\approx -s\pi(0) \int_{\mathbb{R}} g \mathbb{P}_G(dg) \\ &= -\pi(0) \mathbb{E}[G(\mathbf{X})] s. \end{aligned}$$

Therefore, $\Delta_p(s)$ varies linearly with s when the first moment of $G(\mathbf{X})$ exists and is nonzero.

Next, consider the scaling of the cumulant generating function:

$$\begin{aligned} \Delta_K(s, t) &:= K(s, t) - K(0, t) \\ &= \log \mathbb{E} \left[e^{t\bar{F}(G(\mathbf{X});s)} \right] - \frac{t}{2} \\ &= \log \mathbb{E} \left[1 + t \left(\bar{F}(G(\mathbf{X});s) - \frac{1}{2} \right) + O(t^2) \right] \\ &\approx -t \pi(0) \mathbb{E}[G(\mathbf{X})] s. \end{aligned}$$

Thus, for small $|t|$ and finite, nonzero $\mathbb{E}[G(\mathbf{X})]$, the cumulant generating function departs from its small- s fixed point linearly in s .