Log-concave distributions: definitions, properties, and consequences

Jon A. Wellner

University of Washington, Seattle; visiting Heidelberg

Seminaire Point de vue, Université Paris-Diderot Paris 7

23 January 2012

Seminaire Point de vue, Paris

Part 1

Based on joint work with:

- Fadoua Balabdaoui
- Kaspar Rufibach
- Arseni Seregin

Outline, Part 1

- 1: Log-concave densities / distributions: definitions
- 2: Properties of the class
- 3: Some consequences (statistics and probability)
- 4: Strong log-concavity: definitions
- 5: Examples & counterexamples
- 6: Some consequences, strong log-concavity
- 7. Questions & problems

1. Log-concave densities / distributions: definitions

Suppose that a density f can be written as

$$f(x) \equiv f_\varphi(x) = \exp(\varphi(x)) = \exp\left(-(-\varphi(x))\right)$$

where φ is concave (and $-\varphi$ is convex). The class of all densities f on $\mathbb{R}$, or on $\mathbb{R}^d$, of this form is called the class of *log-concave* densities, $\mathcal{P}_{log-concave} \equiv \mathcal{P}_0$.

Note that f is log-concave if and only if :

- $\log f(\lambda x + (1-\lambda)y) \geq \lambda \log f(x) + (1-\lambda)\log f(y)$ for all $0 \leq \lambda \leq 1$ and for all x, y.
- iff $f(\lambda x + (1-\lambda)y) \geq f(x)^\lambda \cdot f(y)^{1-\lambda}$
- iff $f((x+y)/2) \geq \sqrt{f(x)f(y)}$, (assuming f is measurable)
- iff $f((x+y)/2)^2 \geq f(x)f(y)$.

1. Log-concave densities / distributions: definitions

Examples, $\mathbb{R}$

- Example 1: standard normal

$$f(x) = (2\pi)^{-1/2} \exp(-x^2/2),$$
$$-\log f(x) = \frac{1}{2}x^2 + \log\sqrt{2\pi},$$
$$(-\log f)''(x) = 1.$$

- Example 2: Laplace

$$f(x) = 2^{-1} \exp(-|x|),$$
$$-\log f(x) = |x| + \log 2,$$
$$(-\log f)''(x) = 0 \quad \text{for all} \quad x \neq 0.$$

1. Log-concave densities / distributions: definitions

- Example 3: Logistic

$$f(x) = \frac{e^x}{(1+e^x)^2},$$
$$-\log f(x) = -x + 2\log(1+e^x),$$
$$(-\log f)''(x) = \frac{e^x}{(1+e^x)^2} = f(x).$$

- Example 4: Subbotin

$$f(x) = C_r^{-1}\exp(-|x|^r/r), \qquad C_r = 2\Gamma(1/r)r^{1/r-1},$$
$$-\log f(x) = r^{-1}|x|^r + \log C_r,$$
$$(-\log f)''(x) = (r-1)|x|^{r-2}, \quad r \ge 1, \quad x \ne 0.$$

1. Log-concave densities / distributions: definitions

- Many univariate parametric families on $\mathbb{R}$ are log-concave, for example:
  - ▷ Normal (μ, σ^2)
  - ▷ Uniform(a, b)
  - ▷ Gamma(r, λ) for $r \geq 1$
  - ▷ Beta(a, b) for $a, b \geq 1$
  - ▷ Subbotin(r) with $r \geq 1$.
- t_r densities with $r > 0$ are not log-concave
- Tails of log-concave densities are necessarily sub-exponential: i.e. if $X \sim f \in PF_2$, then $E\exp(c|X|) < \infty$ for some $c > 0$.

1. Log-concave densities / distributions: definitions

Log-concave densities on $\mathbb{R}^d$:

- A density f on $\mathbb{R}^d$ is log-concave if $f(x) = \exp(\varphi(x))$ with φ concave.
- Examples
  - ▷ The density f of $X \sim N_d(\mu, \Sigma)$ with Σ positive definite:

$$f(x) = f(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right),$$

$$-\log f(x) = \frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) - (1/2)\log(2\pi|\Sigma),$$

$$D^2(-\log f)(x) \equiv \left(\frac{\partial^2}{\partial x_i \partial x_j}(-\log f)(x), i, j = 1, \ldots, d \right) = \Sigma^{-1}.$$

  - ▷ If $K \subset \mathbb{R}^d$ is compact and convex, then $f(x) = 1_K(x)/\lambda(K)$ is a log-concave density.

1. Log-concave densities / distributions: definitions

Log-concave measures:
Suppose that P is a probability measure on $(\mathbb{R}^d, \mathcal{B}_d)$. P is a log-concave measure if for all nonempty $A, B \in \mathcal{B}_d$ and $\lambda \in (0,1)$ we have

$$P(\lambda A + (1-\lambda)B) \geq \{P(A)\}^{\lambda}\{P(B)\}^{1-\lambda}.$$

- A set $A \subset \mathbb{R}^d$ is *affine* if $tx + (1-t)y \in A$ for all $x, y \in A$, $t \in \mathbb{R}$.
- The *affine hull* of a set $A \subset \mathbb{R}^d$ is the smallest affine set containing A.

Theorem. (Prékopa (1971, 1973), Rinott (1976)). Suppose P is a probability measure on $\mathcal{B}_d$ such that the affine hull of $\text{supp}(P)$ has dimension d. Then P is log-concave if and only if there is a log-concave (density) function f on $\mathbb{R}^d$ such that

$$P(B) = \int_B f(x)dx \quad \text{for all} \quad B \in \mathcal{B}_d.$$

2. Properties of log-concave densities

Properties: log-concave densities on $\mathbb{R}$:

- A density f on $\mathbb{R}$ is log-concave if and only if its convolution with any unimodal density is again unimodal (Ibragimov, 1956).
- Every log-concave density f is unimodal (but need not be symmetric).
- $\mathcal{P}_0$ is closed under convolution.
- $\mathcal{P}_0$ is closed under weak limits

2. Properties of log-concave densities

Properties: log-concave densities on $\mathbb{R}^d$:

- Any log$-$concave f is unimodal.
- The level sets of f are closed convex sets.
- Log-concave densities correspond to log-concave measures. Prékopa, Rinott.
- Marginals of log-concave distributions are log-concave: if $f(x,y)$ is a log-concave density on $\mathbb{R}^{m+n}$, then

$$g(x) = \int_{\mathbb{R}^n} f(x,y)dy$$

  is a log-concave density on $\mathbb{R}^m$. Prékopa, Brascamp-Lieb.
- Products of log-concave densities are log-concave.
- $\mathcal{P}_0$ is closed under convolution.
- $\mathcal{P}_0$ is closed under weak limits.

3. Some consequences and connections (statistics and probability)

- (a) f is log-concave if and only if $\det((f(x_i - y_j))_{i,j\in\{1,2\}}) \ge 0$ for all $x_1 \le x_2$, $y_1 \le y_2$; i.e f is a Polya frequency density of order 2; thus

$$\text{log-concave} \;=\; PF_2 = \text{strongly uni-modal}$$

- (b) The densities $p_\theta(x) \equiv f(x-\theta)$ for $\theta \in \mathbb{R}$ have monotone likelihood ratio (in x) if and only if f is log-concave.

Proof of (b): $p_\theta(x) = f(x-\theta)$ has MLR iff

$$\frac{f(x-\theta')}{f(x-\theta)} \le \frac{f(x'-\theta')}{f(x'-\theta)} \quad \text{for all} \quad x < x', \;\; \theta < \theta'$$

This holds if and only if

$$\log f(x-\theta') + \log f(x'-\theta) \le \log f(x'-\theta') + \log f(x-\theta). \qquad (1)$$

Let $t = (x'-x)/(x'-x+\theta'-\theta)$ and note that

3. Some consequences and connections (statistics and probability)

$$x - \theta = t(x - \theta') + (1 - t)(x' - \theta),$$
$$x' - \theta' = (1 - t)(x - \theta') + t(x' - \theta)$$

Hence log-concavity of f implies that

$$\log f(x - \theta) \geq t \log f(x - \theta') + (1 - t) \log f(x' - \theta),$$
$$\log f(x' - \theta') \geq (1 - t) \log f(x - \theta') + t \log f(x' - \theta).$$

Adding these yields (1); i.e. f log-concave implies $p_\theta(x)$ has MLR in x.

Now suppose that $p_\theta(x)$ has MLR so that (1) holds. In particular that holds if x, x', θ, θ' satisfy $x - \theta' = a < b = x' - \theta$ and $t = (x' - x)/(x' - x + \theta' - \theta) = 1/2$, so that $x - \theta = (a + b)/2 = x' - \theta'$. Then (1) becomes

$$\log f(a) + \log f(b) \leq 2 \log f((a + b)/2).$$

This together with measurability of f implies that f is log-concave.

3. Some consequences and connections (statistics and probability)

Proof of (a): Suppose f is PF_2. Then for $x < x'$, $y < y'$,

$$\det \begin{pmatrix} f(x-y) & f(x-y') \\ f(x'-y) & f(x'-y') \end{pmatrix}$$
$$= f(x-y)f(x'-y') - f(x-y')f(x'-y) \ge 0$$

if and only if

$$f(x-y')f(x'-y) \le f(x-y)f(x'-y'),$$

or, if and only if

$$\frac{f(x-y')}{f(x-y)} \le \frac{f(x'-y')}{f(x'-y)}.$$

That is, $p_y(x)$ has MLR in x. By (b) this is equivalent to f log-concave.

3. Some consequences and connections (statistics and probability)

Theorem. (Brascamp-Lieb, 1976). Suppose $X \sim f = e^{-\varphi}$ with φ convex and $D^2\varphi > 0$, and let $g \in C^1(\mathbb{R}^d)$. Then

$$Var_f(g(X)) \le E\langle (D^2\varphi)^{-1}\nabla g(X), \nabla g(X)\rangle.$$

(Poincaré - type inequality for log-concave densities)

3. Some consequences and connections (statistics and probability)

Further consequences: Peakedness and majorization

Theorem 1. (Proschan, 1965) Suppose that f on $\mathbb{R}$ is log-concave and symmetric about 0. Let $X_1, \ldots, X_n$ be i.i.d. with density f, and suppose that $p, p' \in \mathbb{R}_+^n$ satisfy

- $p_1 \ge p_2 \ge \cdots \ge p_n$, $p'_1 \ge p'_2 \ge \cdots \ge p'_n$,
- $\sum_1^k p'_j \le \sum_1^k p_j$, $k \in \{1, \ldots, n\}$,
- $\sum_1^n p_j = \sum_1^n p'_j = 1$.

(That is, $\underline{p}' \prec \underline{p}$.) Then $\sum_1^n p'_j X_j$ is strictly more peaked than $\sum_1^n p_j X_j$:

$$P\left(|\sum_1^n p'_j X_j| \ge t \right) < P\left(|\sum_1^n p_j X_j| \ge t \right) \quad \text{for all} \quad t \ge 0.$$

3. Some consequences and connections (statistics and probability)

Example: $p_1 = \cdots = p_{n-1} = 1/(n-1)$, $p_n = 0$, while $p'_1 = \cdots = p'_n = 1/n$. Then $\underline{p} \succ \underline{p}'$ (since $\sum_1^n p_j = \sum_1^n p'_j = 1$ and $\sum_1^k p_j = k/(n-1) \ge k/n = \sum_1^k p'_j$), and hence if $X_1, \ldots, X_n$ are i.i.d. f symmetric and log-concave,

$$P(|\overline{X}_n| \ge t) < P(|\overline{X}_{n-1}| \ge t) < \cdots < P(|X_1| \ge t) \quad \text{for all} \quad t \ge 0.$$

Definition: A $d-$dimensional random variable X is said to be *more peaked* than a random variable Y if both X and Y have densities and

$$P\,(Y \in A) \ge P\,(X \in A) \quad \text{for all} \quad A \in \mathcal{A}_d,$$

the class of subsets of $\mathbb{R}^d$ which are compact, convex, and symmetric about the origin.

3. Some consequences and connections (statistics and probability)

Theorem 2. (Olkin and Tong, 1988) Suppose that f on $\mathbb{R}^d$ is log-concave and symmetric about 0. Let $X_1, \ldots, X_n$ be i.i.d. with density f, and suppose that $a, b \in \mathbb{R}^n$ satisfy

- $a_1 \ge a_2 \ge \cdots \ge a_n$, $b_1 \ge b_2 \ge \cdots \ge b_n$,
- $\sum_1^k a_j \le \sum_1^k b_j$, $k \in \{1, \ldots, n\}$,
- $\sum_1^n a_j = \sum_1^n b_j$.

(That is, $\underline{a} \prec \underline{b}$.)

Then $\sum_1^n a_j X_j$ is more peaked than $\sum_1^n b_j X_j$:

$$P\left(\sum_1^n a_j X_j \in A\right) \ge P\left(\sum_1^n b_j X_j \in A\right) \quad \text{for all} \quad A \in \mathcal{A}_d$$

In particular,

$$P\left(\|\sum_1^n a_j X_j\| \ge t\right) \le P\left(\|\sum_1^n b_j X_j\| \ge t\right) \quad \text{for all} \quad t \ge 0.$$

3. Some consequences and connections (statistics and probability)

Corollary: If g is non-decreasing on $\mathbb{R}^+$ with $g(0) = 0$, then

$$Eg\left(\|\sum_1^n a_j X_j\|\right) \le Eg\left(\|\sum_1^n b_j X_j\|\right).$$

Another peakedness result:

Suppose that $\underline{Y} = (Y_1, \ldots, Y_n)$ where $Y_j \sim N(\mu_j, \sigma^2)$ are independent and $\mu_1 \le \ldots \le \mu_n$; i.e. $\underline{\mu} \in K_n$ where $K_n \equiv \{\underline{x} \in \mathbb{R}^n : \ x_1 \le \cdots \le x_n\}$. Let

$$\underline{\widehat{\mu}}_n = \Pi(\underline{Y}|K_n),$$

the least squares projection of $\underline{Y}$ onto K_n. It is well-known that

$$\underline{\widehat{\mu}}_n = \left(\min_{s \ge i} \max_{r \le i} \frac{\sum_{j=r}^s Y_j}{s-r+1}, \ i = 1, \ldots, n\right).$$

3. Some consequences and connections (statistics and probability)

Theorem 3. (Kelly) If $\underline{Y} \sim N_n(\underline{\mu}, \sigma^2 I)$ and $\underline{\mu} \in K_n$, then $\widehat{\mu}_k - \mu_k$ is more peaked than $Y_k - \mu_k$ for each $k \in \{1, \ldots, n\}$; that is

$$P(|\widehat{\mu}_k - \mu_k| \le t) \ge P(|Y_k - \mu_k| \le t) \quad \text{for all} \quad t > 0, \quad k \in \{1, \ldots, n\}.$$

Question: Does Kelly's theorem continue to hold if the normal distribution is replaced by an arbitrary log-concave joint density symmetric about $\underline{\mu}$?

4. Strong log-concavity: definitions

Definition 1. A density f on $\mathbb{R}$ is *strongly log-concave* if

$$f(x) = h(x) c\phi(cx) \quad \text{for some} \quad c > 0$$

where h is log-concave and $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$.
Sufficient condition: $\log f \in C^2(\mathbb{R})$ with $(-\log f)''(x) \geq c^2 > 0$ for all x.

Definition 2. A density f on $\mathbb{R}^d$ is *strongly log-concave* if

$$f(x) = h(x) c\gamma(cx) \quad \text{for some} \quad c > 0$$

where h is log-concave and γ is the $N_d(0, cI_d)$ density.
Sufficient condition: $\log f \in C^2(\mathbb{R}^d)$ with $D^2(-\log f)(x) \geq c^2 I_d$ for some $c > 0$ for all $x \in \mathbb{R}^d$.

These agree with *strong convexity* as defined by Rockafellar & Wets (1998), p. 565.

5. Examples & conterexamples

Examples

Example 1. $f(x) = h(x)\phi(x)/\int h\phi dx$ where h is the logistic density, $h(x) = e^x/(1+e^x)^2$.

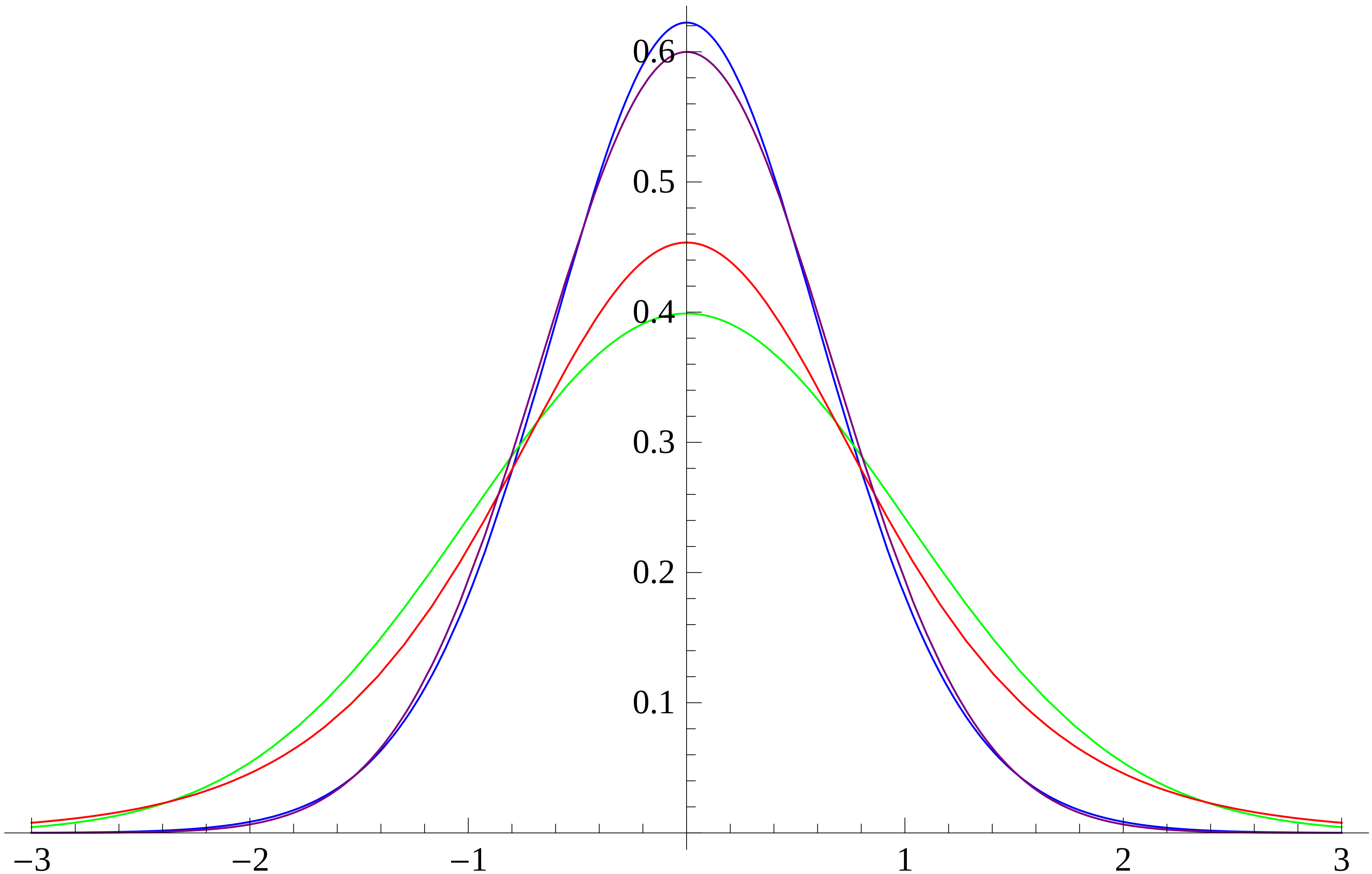
Example 2. $f(x) = h(x)\phi(x)/\int h\phi dx$ where h is the Gumbel density. $h(x) = \exp(x - e^x)$.

Example 3. $f(x) = h(x)h(-x)/\int h(y)h(-y)dy$ where h is the Gumbel density.

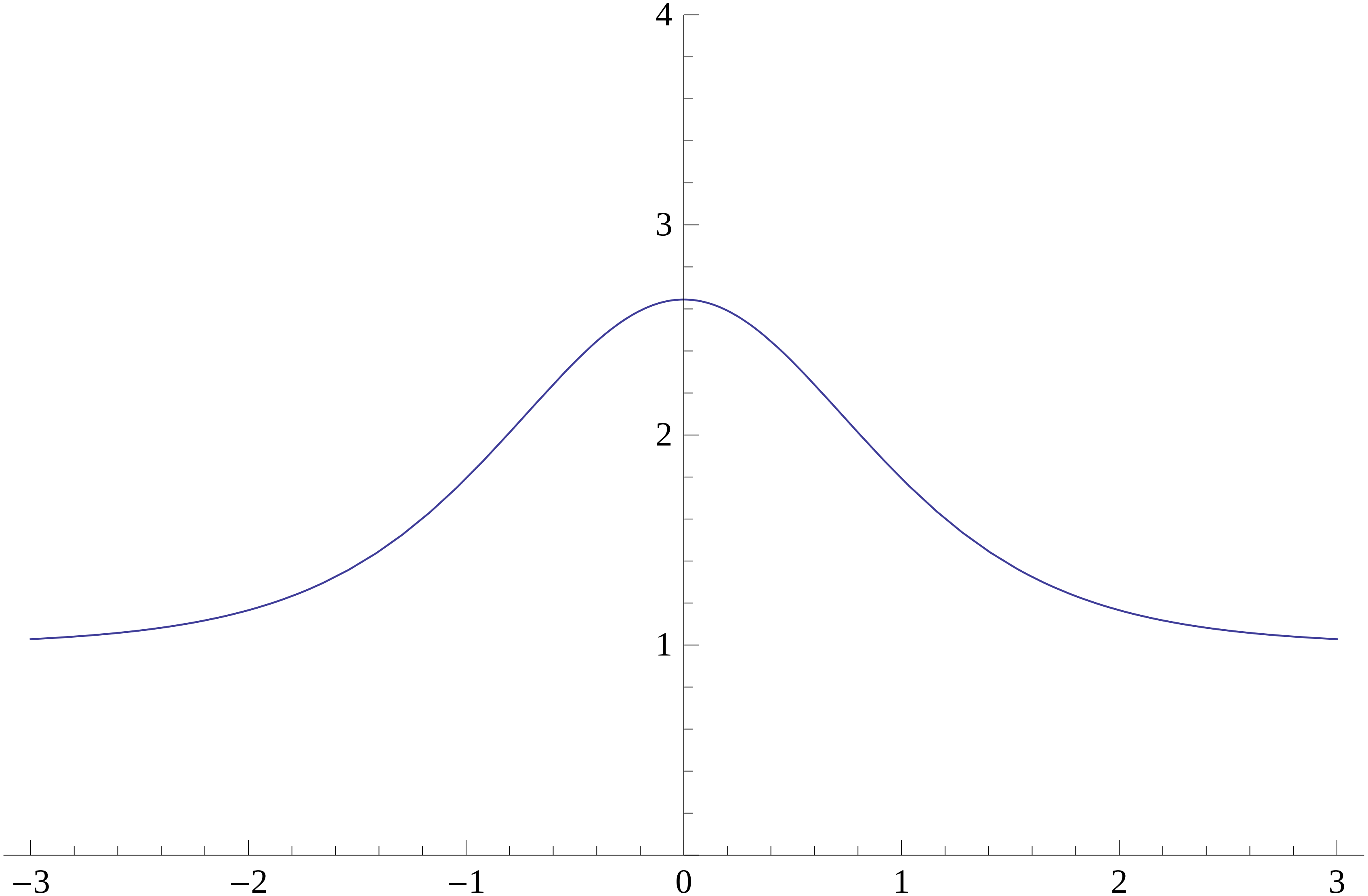
Counterexamples

Counterexample 1. f logistic: $f(x) = e^x/(1+e^x)^2$; $(-\log f)''(x) = f(x)$.

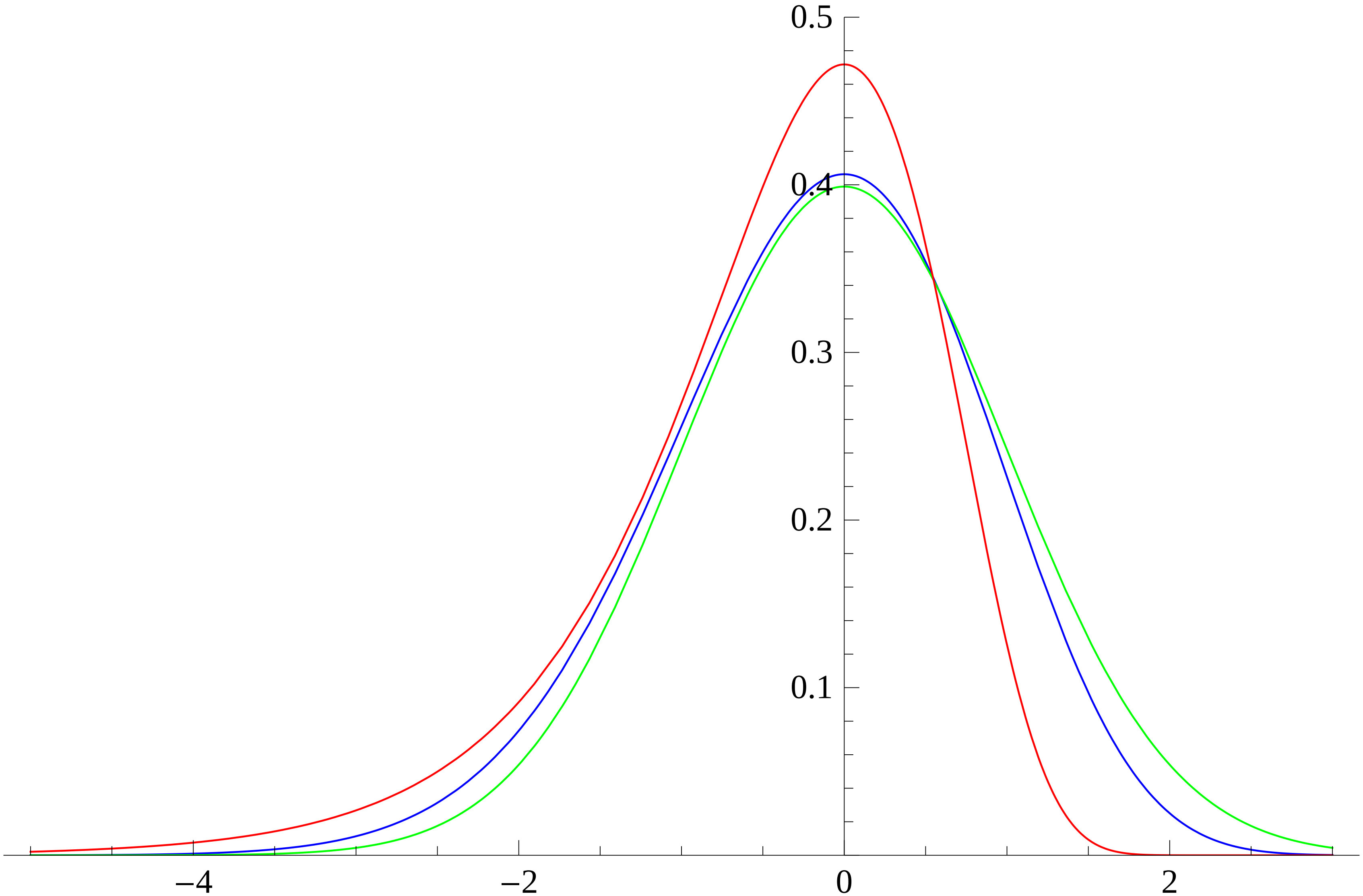
Counterexample 2. f Subbotin, $r \in [1,2) \cup (2,\infty)$; $f(x) = C_r^{-1}\exp(-|x|^r/r)$; $(-\log f)''(x) = (r-2)|x|^{r-2}$.



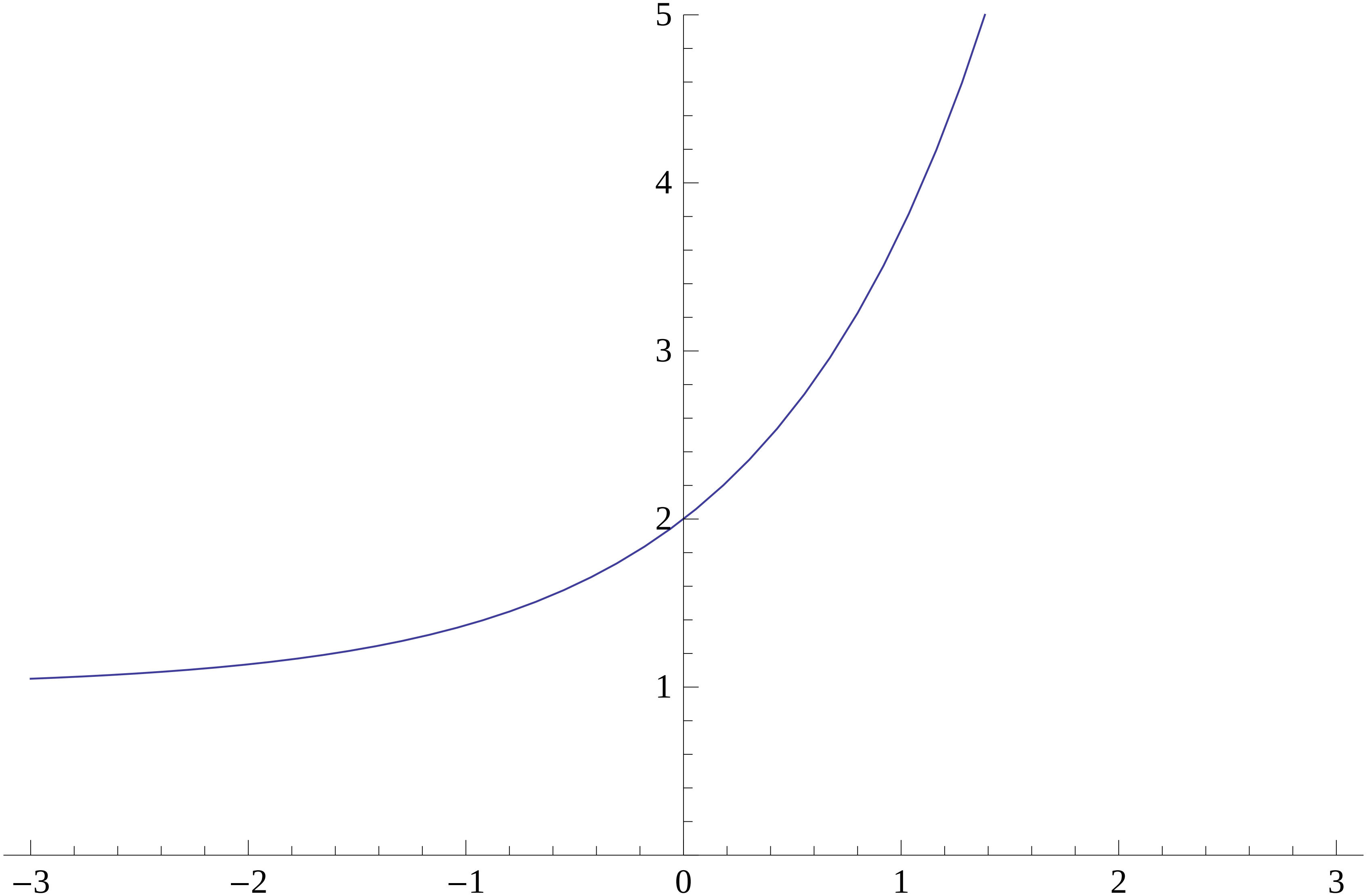
Ex. 1: Logistic (red) perturbation of $N(0, 1)$ (green): f (blue)



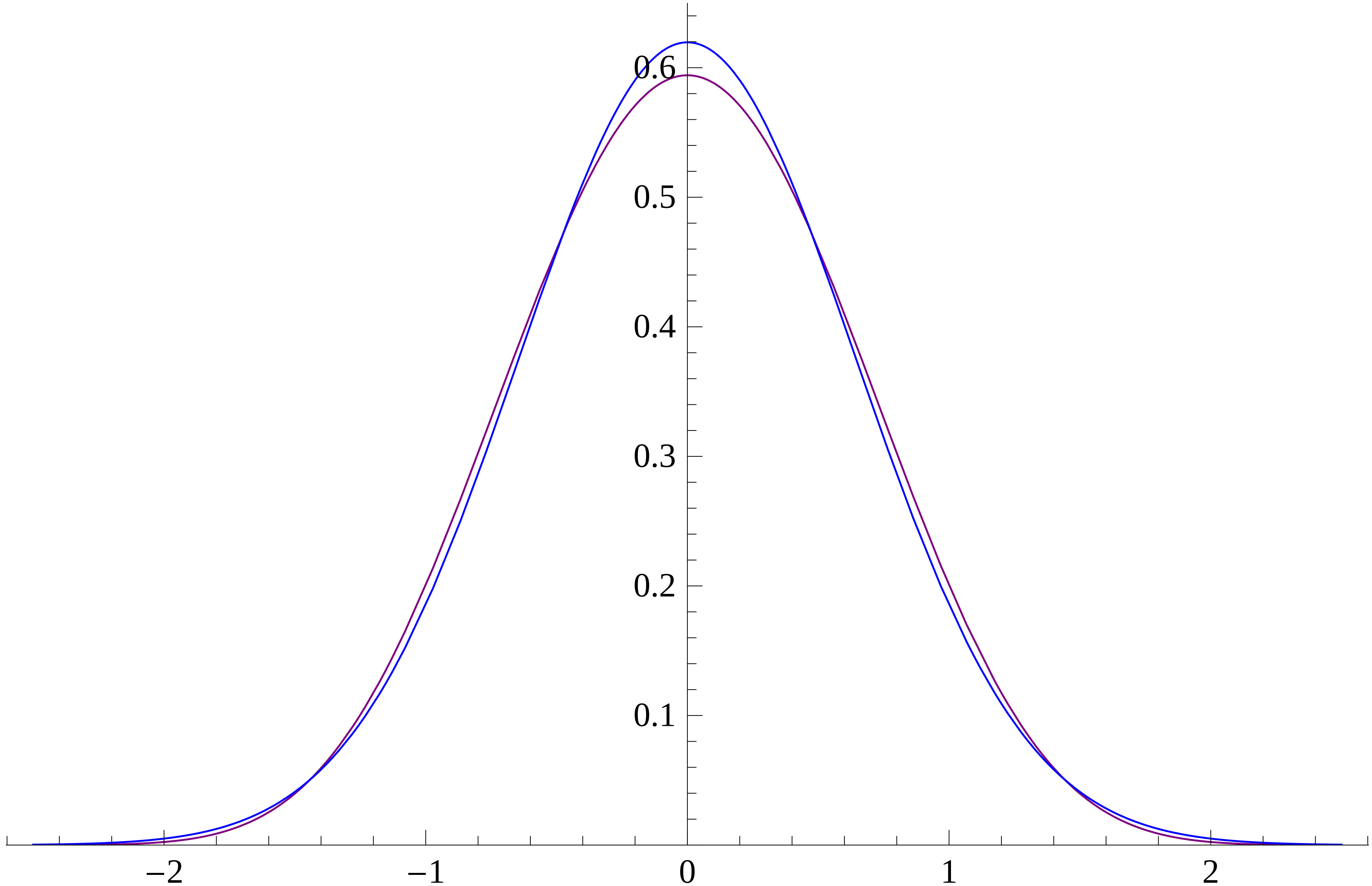
Ex. 1: $(-\log f)''$, Logistic perturbation of $N(0,1)$



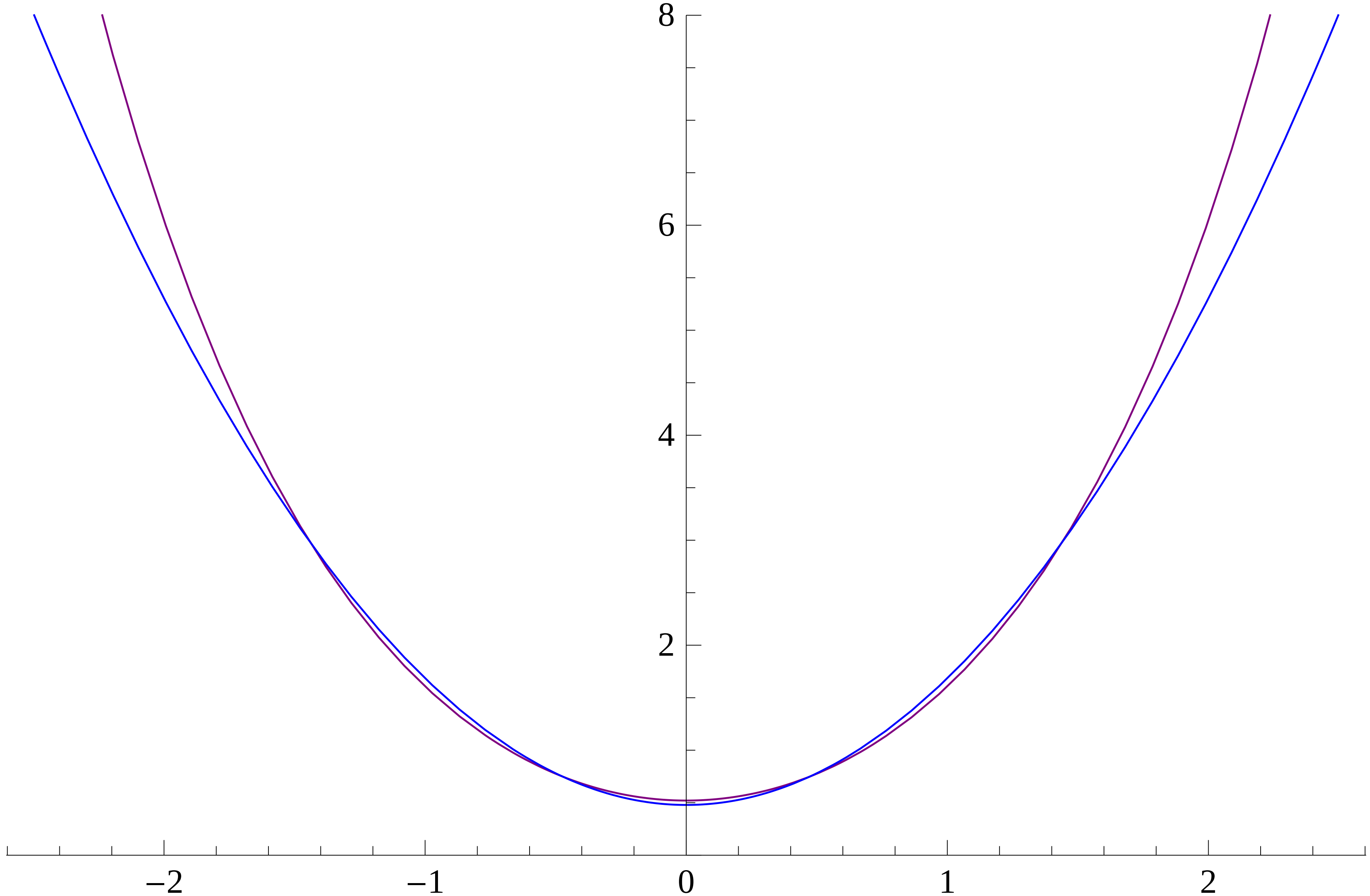
Ex. 2: Gumbel (red) perturbation of $N(0,1)$ (green): f (blue)



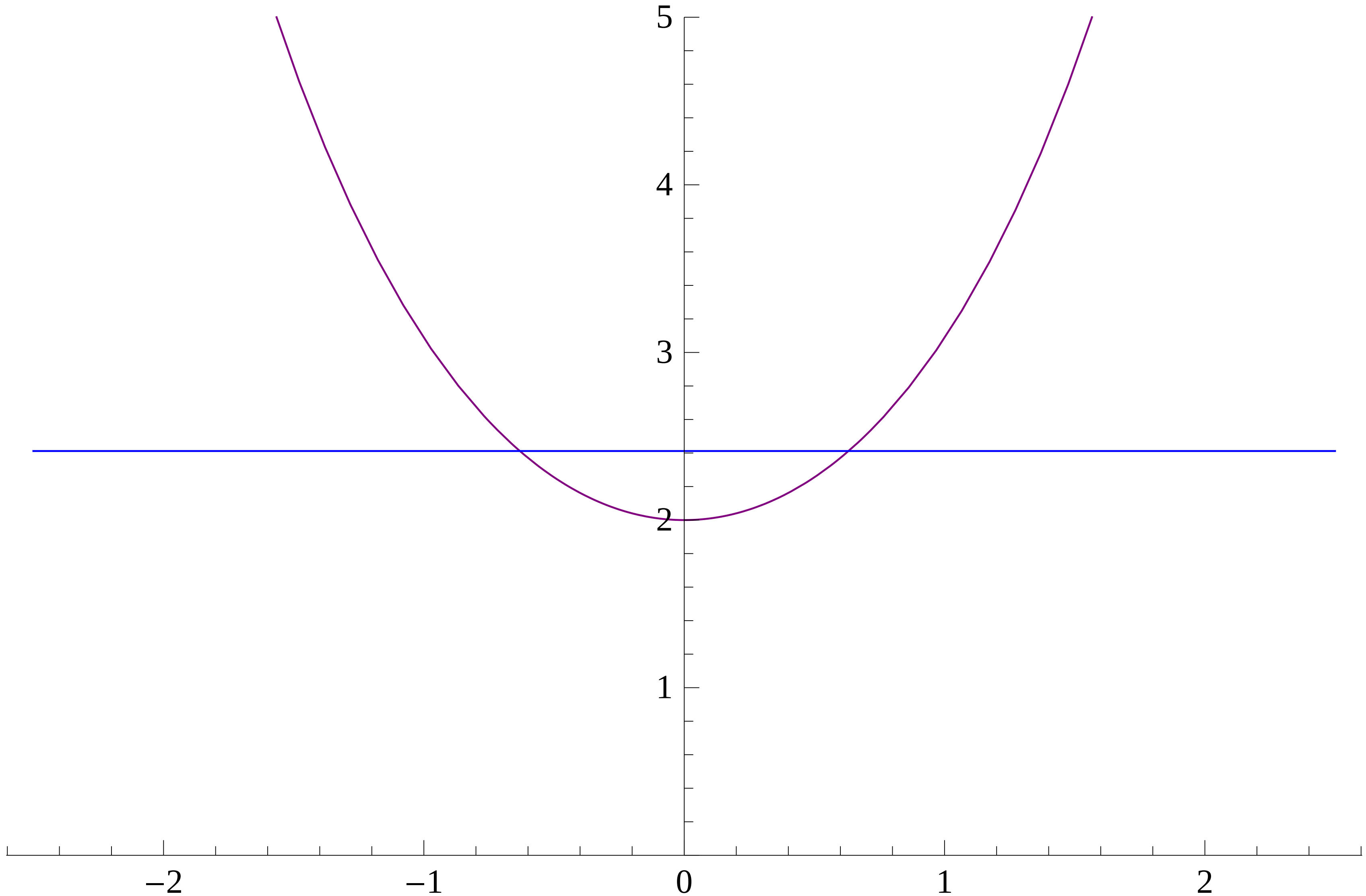
Ex. 2: $(-\log f)''$, Gumbel perturbation of $N(0,1)$



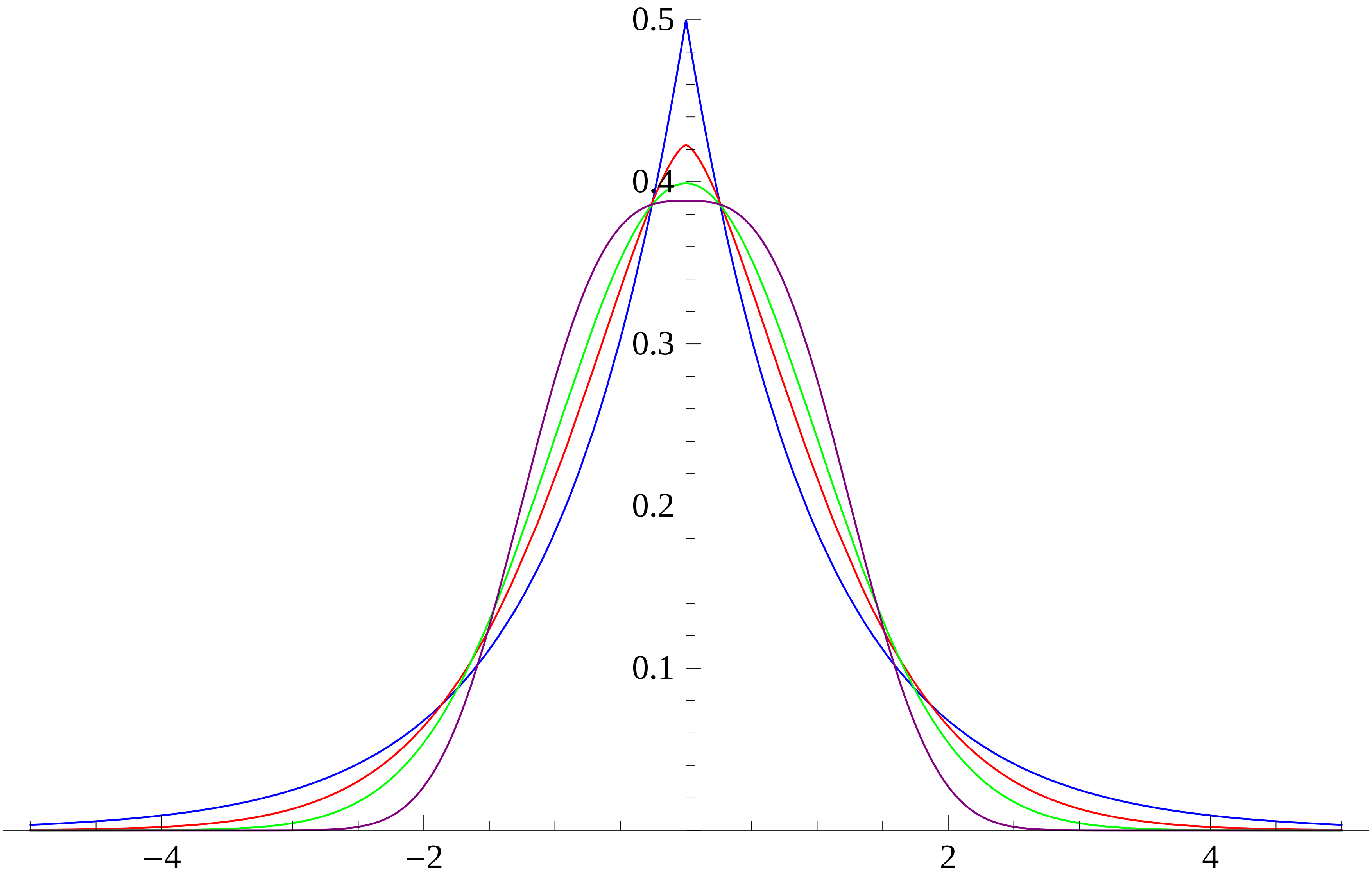
Ex. 3: Gumbel $(\cdot) \times$ Gumbel$(-\cdot)$ (purple); $N(0, V_f)$ (blue)



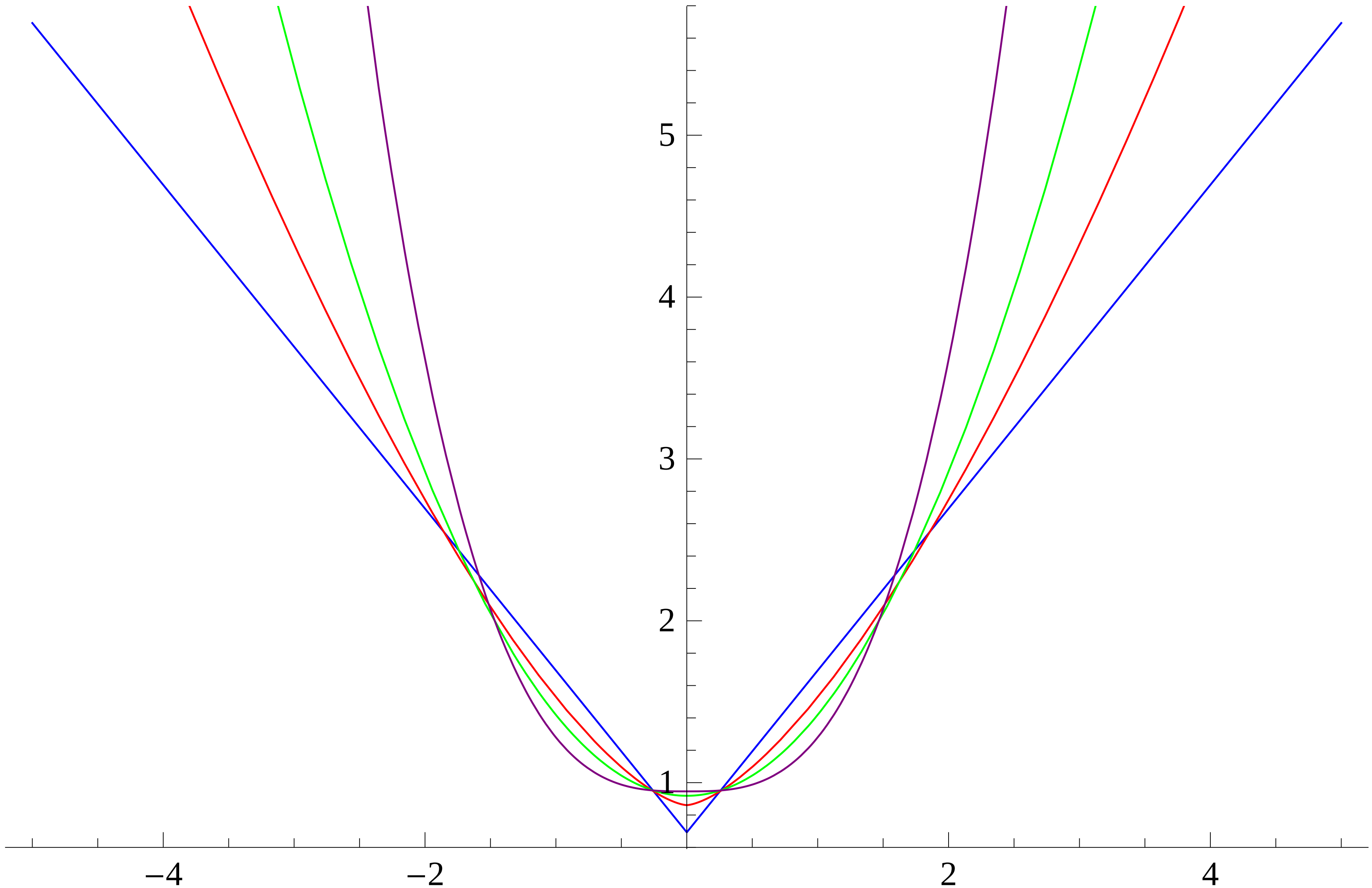
Ex. 3: $-\log$Gumbel$(\cdot) \times$ Gumbel$(-\cdot)$ (purple); $-\log N(0, V_f)$ (blue)



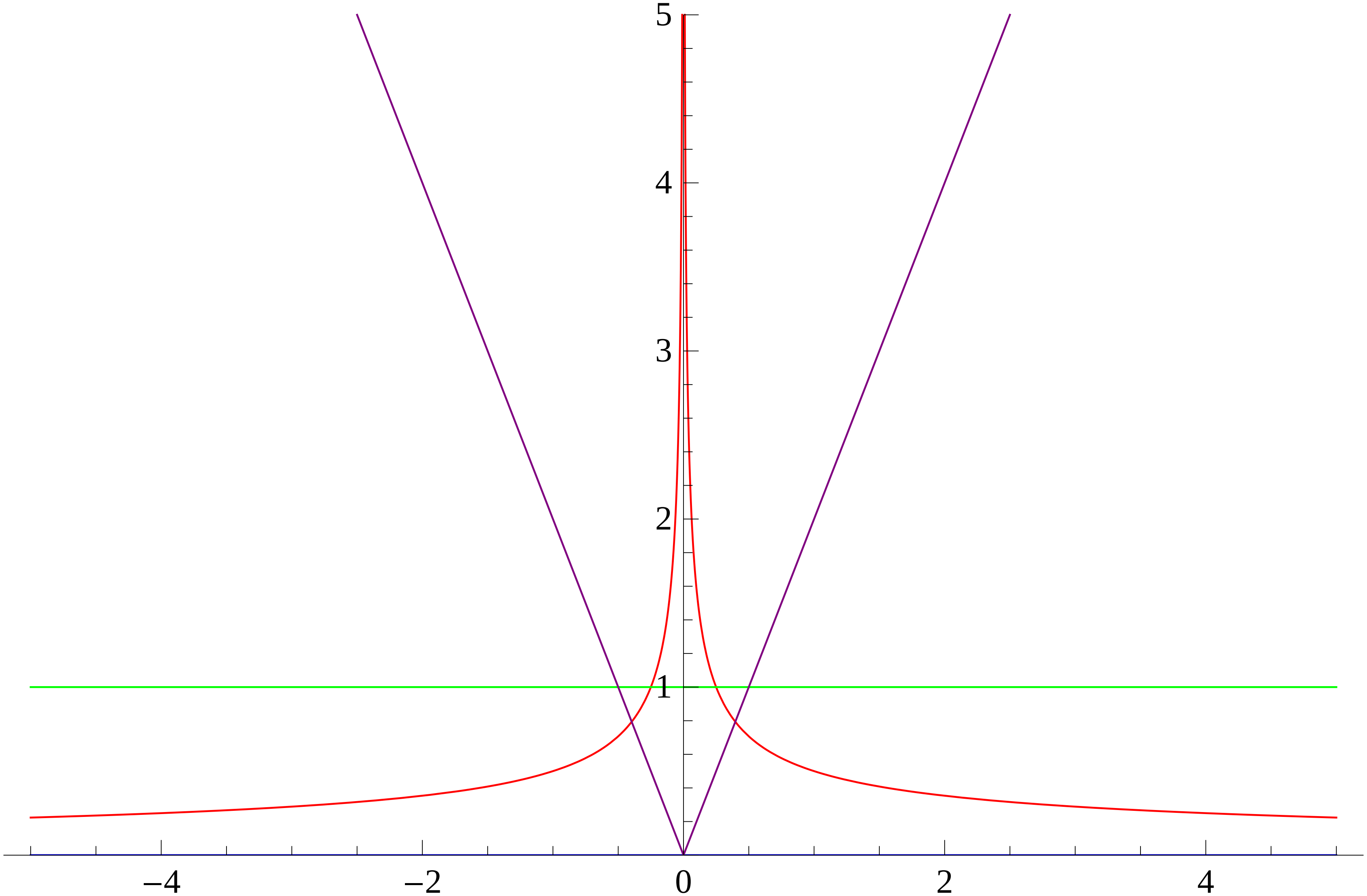
Ex. 3: $D^2(-\log \text{Gumbel}(\cdot) \times \text{Gumbel}(-\cdot))$ (purple); $D^2(-\log N(0, V_f))$ (blue)



Subbotin f_r $r = 1$ (blue), $r = 1.5$ (red), $r = 2$ (green), $r = 3$ (purple)



$-\log f_r$: $r = 1$ (blue), $r = 1.5$ (red), $r = 2$ (green), $r = 3$ (purple)



$(-\log f_r)''$: $r = 1$ (blue), $r = 1.5$ (red), $r = 2$ (green), $r = 3$ (purple)

6. Some consequences, strong log-concavity

First consequence

Theorem. (Hargé, 2004). Suppose $X \sim N_n(\mu, \boldsymbol{\Sigma})$ with density γ and Y has density $h \cdot \gamma$ with h log-concave, and let $g : \mathbb{R}^n \to \mathbb{R}$ be convex. Then

$$Eg(Y - E(Y)) \le Eg(X - EX)).$$

Equivalently, with $\mu = EX$, $\nu = EY = E(Xh(X))/Eh(X)$, and $\tilde{g} \equiv g(\cdot + \mu)$

$$E\{\tilde{g}(X - \nu + \mu)h(X)\} \le E\tilde{g}(X) \cdot Eh(X).$$

6. Some consequences, strong log-concavity

More consequences

Corollary. (Brascamp-Lieb, 1976). Suppose $X \sim f = \exp(-\varphi)$ with $D^2\varphi \geq \lambda I_d$, $\lambda > 0$, and let $g \in C^1(\mathbb{R}^d)$. Then

$$Var_f(g(X)) \leq E\langle (D^2\varphi)^{-1}\nabla g(X), \nabla g(X)\rangle \leq \frac{1}{\lambda} E|\nabla g(X)|^2.$$

(Poincaré inequality for strongly log-concave densities; improvements by Hargé (2008))

Theorem. (Caffarelli, 2002). Suppose $X \sim N_d(0, I)$ with density γ_d and Y has density $e^{-v} \cdot \gamma_d$ with v convex. Let $T = \nabla\varphi$ be the unique gradient of a convex map φ such that $\nabla\varphi(X) \stackrel{d}{=} Y$. Then

$$0 \leq D^2\varphi \leq I_d.$$

(cf. Villani (2003), pages 290-291)

7. Questions & problems

- Does strong log-concavity occur *naturally*? Are there natural examples?

- Are there large classes of strongly log-concave densities in connection with other known classes such as PF_∞ (Pólya frequency functions of order infinity) or L. Bondesson's class HM_∞ of completely hyperbolically monotone densities?

- Does Kelly's peakedness result for projection onto the ordered cone K_n continue to hold with Gaussian replaced by log-concave (or symmetric log concave)?

Selected references:

- Villani, C. (2002). *Topics in Optimal Transportation*. Amer. Math Soc., Providence.
- Dharmadhikari, S. and Joag-dev, K. (1988). *Unimodality, Convexity, and Applications*. Academic Press.
- Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press.
- Marshall, A. W., Olkin, I., and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*. Second edition, Springer.
- Rockfellar, R. T. and Wets, R. (1998). *Variational Analysis*. Springer.
- Hargé, G. (2004). A convex/log-concave correlation inequality for Gaussian measure and an application to abstract Wiener spaces. *Probab. Theory Related Fields* **130**, *415-440.*
- *Brascamp, H. J. and Lieb, E. H. (1976). On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. J. Functional Analysis* **22**, *366-389.*
- *Hargé, G. (2008). Reinforcement of an inequality due to Brascamp and Lieb. J. Funct. Anal.,* **254, 267-300.**