Chernoff’s distribution is log-concave
But why? (And why does it matter?)

Jon A. Wellner

University of Washington, Seattle

University of Michigan

April 1, 2011
Woodrofe Seminar

Based on joint work with: Fadoua Balabdaoui
Outline

• 1: Background: how does Chernoff’s distribution appear?
• 2: Groeneboom’s formula and graphical evidence for log-concavity
• 3: Log-concavity: a brief review and some consequences
• 4: An indirect approach to log-concavity of $f_Z$.
• 5: Summary; further questions and open problems.
1. Background: how does Chernoff’s distribution appear?

- Nonparametric estimation of a monotone function!
- Three problems
  - Estimation of an increasing regression function: Ayer, Brunk, Ewing, Reid, Silverman (1955), van Eeden (1957)
1. Background:
1. Background:

Let $X_1, \cdots, X_n$ be independent and identically distributed with common density $f$ which is unimodal with known mode $\mu$. The author considers the asymptotic behavior $\hat{f}_n(x) - f(x)$, where $\hat{f}_n$ denotes the maximum likelihood estimate of $f$. The author shows that if $f$ is differentiable at $x$, $f'(x) \neq 0 \neq f(x)$, then $c_n(\hat{f}(x) - f(x))$ has a non-degenerate asymptotic distribution where

$$c_n = \left(\frac{2n}{f(x)|f'(x)|}\right)^{1/3}.$$

Moreover, the density of the asymptotic distribution is given in terms of the partial derivatives of $\mu$, where $\mu(x, z) \Pr(W(t) > t^2, \text{ for some } t > z|W(z) = x)$.

Reviewed by M. B. Woodroofe

© Copyright American Mathematical Society 1971, 2011
1. Background:

- In each case:
  - There is a monotone function $m$ to be estimated
  - There is a natural nonparametric estimator $\hat{m}_n$.
  - If $m'(x_0) \neq 0$ and $m'$ continuous at $x_0$, then
    \[
    n^{1/3}(\hat{m}_n(x_0) - m(x_0)) \to_d C(m, x_0) Z
    \]
    where
    \[
    Z = \arg\max \{ W(t) - t^2 \}
    \]
    $\overset{d}{=} \frac{1}{2}$ slope at zero of least concave majorant of $W(t) - t^2$. }
1. Background:

First appearance of $Z$:
Chernoff (1964), *Estimation of the mode*:

- $X_1, \ldots, X_n$ i.i.d. with density $f$ and distribution function $F$.
- Fix $a > 0$; $\hat{x}_a$ $\equiv$ center of the interval of length $2a$ containing the most observations.
- $x_a \equiv$ center of the interval of length $2a$ maximizing $F(x + a) - F(x - a)$.
- Chernoff shows:
  
  $n^{1/3}(\hat{x}_a - x_a) \xrightarrow{d} \left(\frac{8f(x_a+a)}{c}\right)^{1/3} Z$ where
  
  $c \equiv f'(x_a - a) - f'(x_a + a)$.
1. Background:

\[ f_Z(z) = \frac{1}{2}g(z)g(-z) \text{ where} \]

\[ g(t) \equiv \lim_{x \to t^2} \frac{\partial}{\partial x} u(t, x) = \lim_{x \to t^2} u_x(t, x), \]

\[ u(t, x) \equiv P^{(t,x)}(W(z) > z^2, \text{ for some } z \geq t) \text{ is a solution} \]

\[ \frac{\partial}{\partial t} u(t, x) = -\frac{1}{2} \frac{\partial^2}{\partial x^2} u(t, x) \]

under the boundary conditions

\[ u(t, t^2) = \lim_{x \to t^2} u(t, x) = 1, \quad \lim_{x \to -\infty} u(t, x) = 0. \]
1. Background:

Herman Chernoff
2. Groeneboom's formula and graphical evidence for log-concavity

Groeneboom (1985, 1989) analyzed the process

\[ V_c(a) \equiv \sup\{t \in R : W(t) - c(t - a)^2 \text{ is maximal}\}, \]

and along the way showed that \( Z_c \equiv V_c(0) \) has density

\[ f_{Z_c}(t) = \frac{1}{2} g_c(t) g_c(-t) \]

where \( g_c \) has Fourier transform given by

\[ \hat{g}_c(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda s} g_c(s) ds = \frac{2^{1/3} c^{-1/3}}{Ai(i(2c^2)^{-1/3}\lambda)}; \]

here \( Ai \equiv u \) is the Airy function satisfying \( u''(z) - zu(z) = 0 \), and Chernoff’s \( Z \equiv Z_1 \) and \( g \equiv g_1 \). Thus

\[ g_c(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \frac{2^{1/3} c^{-1/3}}{Ai(i(2c^2)^{-1/3}u)} du. \]
2. Groeneboom’s formula

Asymptotics of $g_c$ and $f_Z$; Groeneboom (1989): as $t \to \infty$,

$$g_c(t) \sim 4ct \exp \left( -\frac{2}{3}c^2 t^3 \right), \quad \text{and}$$

$$f_Z(t) \sim \frac{(1/2)(4c)^{4/3}}{Ai'(a_1)} |t| \exp \left( -\frac{2}{3}c^2 |t|^3 + (2c^2)^{1/3}a_1 |t| \right)$$

where $a_1 \approx -2.3381$ is the largest zero of the Airy function $Ai$ and where $Ai'(a_1) \approx 0.7022$. 
2. Groeneboom’s formula

George Biddell Airy, 1801-1892
2. Groeneboom’s formula

Piet Groeneboom, 1941-1990, Günzburg
2. Groeneboom’s formula

The density $f_Z$
2. Groeneboom’s formula
2. Groeneboom’s formula ...

The function $g$
2. Groeneboom’s formula ...
If \( h(x) \equiv \log g(x) \) and \( h_Z(x) = \log f_Z(x) \), then

\[
\begin{align*}
  h'(x) & = \frac{g'(x)}{g(x)}, \\
  h''(x) & = \frac{g''(x)g(x) - g'(x)^2}{g(x)^2}, \\
  h'_Z(x) & = \frac{g'(x)}{g(x)} - \frac{g'(-x)}{g(-x)}, \\
  h''_Z(x) & = \frac{g''(x)g(-x) + g''(-x)g(x)}{g(x)g(-x)} - \frac{g'(x)^2g(-x)^2 + g'(-x)^2g(x)^2}{g(x)^2g(-x)^2}.
\end{align*}
\]

Thus log concavity of \( g \) holds if the numerator of \( h'' \) is negative, and log concavity of \( f_Z \) holds if the numerator of \( h''_Z \) is negative. Here are plots of these functions.
2. Groeneboom’s formula ...
2. Groeneboom’s formula ...

Numerator, second derivative, $\log f_Z$
2. ... and graphical evidence for log-concavity

Recall that $f$ is log-concave if and only if, for all $x, y \in \mathbb{R}$, $\lambda \in [0, 1],$

$$\log f(\lambda x + (1 - \lambda)y) \geq \lambda \log f(x) + (1 - \lambda) f(y)$$

iff $$f(\lambda x + (1 - \lambda)y) \geq f(x)^{\lambda} f(y)^{1 - \lambda}$$

iff $$f \left( \frac{x + y}{2} \right)^2 \geq f(x)f(y) \quad (\text{Sierpinski, Jensen})$$

iff $$f \left( \frac{x + y}{2} \right)^2 - f(x)f(y) \geq 0$$

iff $$D_f(t, s) \equiv f(t)^2 - f(t + s)f(t - s) \geq 0 \quad \text{for all } t, s.$$ 

The plots on the following pages show the function $D_f(t, s)$ for $f = \phi$, the standard normal density and for $f = f_Z$, Chernoff's density.
The function $D_f(t,s)$, $f = \phi$, standard normal density
2. ... and graphical evidence for log-concavity

The function $D_f(t,s)$, $f = f_Z$, Chernoff’s density
3: Log-concavity: a review and some consequences

Suppose that

\[ f(x) \equiv f_\varphi(x) = \exp(\varphi(x)) = \exp(-(-\varphi(x))) \]

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

Properties of log-concave densities:

- 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.
3. Log-concavity: a review ...
Many parametric families are log-concave, for example:

- Normal \((\mu, \sigma^2)\)
- Uniform\((a, b)\)
- Gamma\((r, \lambda)\) for \(r \geq 1\)
- Beta\((a, b)\) for \(a, b \geq 1\)

- \(t_r\) densities with \(r > 0\) are not log-concave

- Tails of log-concave densities are necessarily sub-exponential

\(P_{\text{log-concave}} = \) the class of “Polyá frequency functions of order 2”, \(PFF_2\), in the terminology of Schoenberg (1951) and Karlin (1968). See Marshall and Olkin (1979), chapter 18, and Dharmadhikari and Joag-Dev (1988), page 150, for nice introductions.
Log-concave densities on $\mathbb{R}^d$:

- A density $f$ on $\mathbb{R}^d$ is log-concave if $f(x) = \exp(\varphi(x))$ with $\varphi$ concave.

- Some properties:

  - Any log–concave $f$ is unimodal
  - The level sets of $f$ are closed convex sets
  - Convolutions of log-concave distributions are log-concave.
  - Marginals of log-concave distributions are log-concave (Brascamp-Lieb).
  - Log-concave densities correspond to log-concave measures. (Prekopa; Borell-Brascamp-Lieb).

$$P(\lambda A + (1-\lambda)B) \geq P(A)^\lambda P(B)^{1-\lambda} \text{ for all } A, B, \lambda \in [0, 1].$$
3. Log-concavity: ... some consequences

**Inequality 1.** Brascamp-Lieb (1976); Hargé (2004): Let $f$ be log-concave and let $g$ be convex on $\mathbb{R}^n$. Suppose that $Y \sim \frac{fd\gamma}{\int fd\gamma}$ where $\gamma$ is the probability measure corresponding to $X \sim N_n(\mu, \Sigma)$ for some $\mu$, $\Sigma$. Then

$$Eg(Y - E(Y)) \leq Eg(X - E(X)).$$

(Brascamp-Lieb for $g(x) = |x|^\alpha$ with $\alpha \geq 1$; Hargé for a general convex function $g$.) Note that if $\gamma$ has density $h$ with respect to Lebesgue measure then $Y$ has density $f \cdot h/\int fh dx$ which is log-concave.

• A 1–dimensional random variable $Y$ is said to be more peaked about $\mu$ than a random variable $X$ about $\nu$ if

$$P(|Y - \mu| \leq t) \geq P(|X - \nu| \leq t) \text{ for all } t.$$  

When $\mu = \nu = 0$ we say that $Y$ is more peaked than $X$ and write $Y \overset{p}{>} X$.

• An $n$–dimensional random variable $Y$ is said to be more peaked than a vector $X$ if they have densities and if

$$P(Y \in A) \geq P(X \in A) \text{ for all } A \in \mathcal{A}_n$$

$\mathcal{A}_n \equiv \{\text{all compact, convex, symmetric (about 0) Borel sets}\}$.

Again we will write $Y \overset{p}{>} X$.  

3. Log-concavity: ... some consequences

Proposition 1. (Sherman, 1955) Suppose that \( X_j \sim f_j \) and \( Y_j \sim g_j \) for \( j = 1, 2 \) are independent random vectors in \( \mathbb{R}^n \) and that the densities \( f_1, f_2, g_1, g_2 \) are bounded, symmetric about the origin, and log-concave. If \( Y_1 \overset{p}{>} X_1 \) and \( Y_2 \overset{p}{>} X_2 \), then

\[
Y_1 + Y_2 \overset{p}{>} X_1 + X_2.
\]

Proposition 2. (Sherman, 1955) Suppose that \( Y_1, \ldots, Y_n \) are independent with bounded densities which are symmetric about the origin and log-concave and similarly for \( X_1, \ldots, X_n \), then \( Y_j \overset{p}{>} X_j \) for \( j \in \{1, \ldots, n\} \) implies

\[
\sum_{j=1}^{n} c_j Y_j \overset{p}{>} \sum_{j=1}^{n} c_j X_j
\]

for all real numbers \( c_j \).
3. Log-concavity: ... some consequences

Proposition 3. (Proschan, 1965) Suppose that $Z_1, \ldots, Z_n$ are i.i.d random variables with log-concave density symmetric about zero. Then if $a, b \in \mathbb{R}^n_+$ with $a \succ b$ (a majorizes b; i.e. $b \in \text{conv}\{\pi a : \pi \in \Pi\}$), then

$$\sum_{j=1}^{n} b_j Z_j \overset{p}{>} \sum_{j=1}^{n} a_j Z_j \quad \text{in } \mathbb{R}.$$ 

Proposition 4. (Olkin and Tong, 1988) Suppose that $Z_1, \ldots, Z_n$ are i.i.d random vectors in $\mathbb{R}^d$ with log-concave density symmetric about zero. Then if $a, b \in \mathbb{R}^n$ with $a \succ b$ (a majorizes b; i.e. $b \in \text{conv}\{\pi a : \pi \in \Pi\}$), then

$$\sum_{j=1}^{n} b_j Z_j \overset{p}{>} \sum_{j=1}^{n} a_j Z_j \quad \text{in } \mathbb{R}^d.$$
Proposition 5. (Kelly, 1989) Suppose that $Y = (Y_1, \ldots, Y_n)$ where $Y_j \sim N(\mu_j, \sigma^2)$ and independent and $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$. Let $\hat{\mu} \equiv TY$ where $T$ is the least-squares projection (or isotonization operator) from $\mathbb{R}^n$ to the ordered cone

$$K_n \equiv \mathbb{R}^n_{\leq} \equiv \{x \in \mathbb{R}^n : x_1 \leq x_2 \leq \ldots \leq x_n\}.$$ 

Then $\hat{\mu}_k - \mu_k \overset{p}{>} Y_k - \mu_k$ for each $k \in \{1, \ldots, n\}$; i.e.

$$P(|\hat{\mu}_k - \mu_k| \leq t) \geq P(|Y_k - \mu_k| \leq t) \text{ for all } t > 0, \ k \in \{1, \ldots, n\}.$$

**Key fact:** (Dharmadhikari and Joag-Dev, 1988) Suppose that $F_n$ has support $(-\infty, \infty)$ and has a log-concave density $f_n$ for each $n \in \mathbb{N}$. Suppose that $F_n \xrightarrow{d} F$ where $F$ is continuous. Then $F$ has a log-concave density $f$.

This means that if we can find some sequence of random variables $Z_n$ such that $Z_n$ has log-concave density $f_n$ for each $n$, and satisfying $Z_n \xrightarrow{d} Z$, then the density $f_Z$ of $Z$ is log-concave.

Thus we consider the simplest possible version of the monotone regression problem as treated in Barlow, Bartholomew, Bremner, and Brunk (1972) (or Robertson, Wright, and Dykstra (1988). Suppose that $r(x) = x$ for $x \in [0, 1]$, and that

$$Y_i = r(x_i) + \epsilon_i \equiv \mu_i + \epsilon_i, \quad i \in \{1, \ldots, n\}$$

where $x_i \equiv i/(n + 1)$ for $1 \leq i \leq n$ and where $\epsilon_i \sim N(0, 1)$. Thus $Y \sim N_n(\mu, I)$. 

The least-squares isotonic estimator is given at $x_0 \in (0, 1)$ by

$$TY = \hat{r}_n(x_0) = \max_{i: x_i \leq x_0} \min_{k: x_k \geq x_0} \frac{\sum_{j=i}^{k} Y_j}{k - i + 1}.$$

It is known from Brunk (1970) (or Wright (1981), Leurgans (1982)) that

$$Z_n \equiv n^{1/3}(\hat{r}_n(x_0) - r(x_0)) \rightarrow_d (r'(x_0)/2)^{1/3}(2Z).$$

Thus our current approach is to show that $\hat{r}_n(x_0)$ in this setting has a log-concave density and hence that $Z_n$ has a log-concave density (for each $n$).
Let $Y \sim N_n(\mu, I)$ where $\mu \in K_n \equiv \mathbb{R}^n_+$, and write

$$\hat{\mu}_n = TY.$$

**Proposition 1.** The vector $\hat{\mu}_n$ has an (induced) distribution on $K_n$ which is a mixture of log-concave measures on $K_n^\circ$ and the $n(n-1)/2$ lower-dimensional faces of $K_n$. The components of the mixed log-concave measure all have log-concave densities with respect to Lebesgue measure on the corresponding subspaces.

General projection plot, $\lambda A + (1 - \lambda)B$

**Proposition 2 (Conjectured).** The one-dimensional marginal distributions of $\hat{\mu}_n$ are all log-concave.

**Proof?!**
5. Summary; further questions and open problems.

Summary:

- Groeneboom’s analysis of $f_Z$ leads to numerical computations strongly suggesting that $f_Z$ is log-concave.
- No analytic proof of log-concavity of $f_Z$ based on Groeneboom’s formula (yet).
- Log-concavity would be a valuable tool in conjunction with a generalization of Kelly’s peaked-ness result for the Gaussian case.
- Current favored proof route: via limit theory and marginal log-concavity of the ordered cone projection map in the Gaussian monotone regression problem.
5. ... further questions and open problems.

- Conjecture 1: Propositions 1 and 2 continue to hold if the normal assumption on the $\epsilon_i$'s is replaced by a log-concave assumption.

- Conjecture 2: (suggested by the asymptotic behavior of $f_Z$ at infinity)

$$f_Z(t) = h(t)\phi(t/b)/b$$

with $h$ log-concave for all $b \geq$ some $b_0$, and hence the Brascamp-Lieb / Hargé inequality yields

$$Eg(Z) \leq Eg(N(0,b^2))$$

for all $g$ convex, $b \geq b_0$.

(I conjecture $b_0 \approx .54$ . . . .)

- Conjecture 3: The indirect approach sketched above will also work for the cases in which “monotone” is replaced by “convex”.

- Conjecture 4: Kelly’s theorem remains valid with the Gaussian hypothesis replaced by a log-concave hypothesis.

- Conjecture 5: Chernoff’s distribution is log-concave!