Likelihood Tilting

Christopher G. Small
University of Waterloo

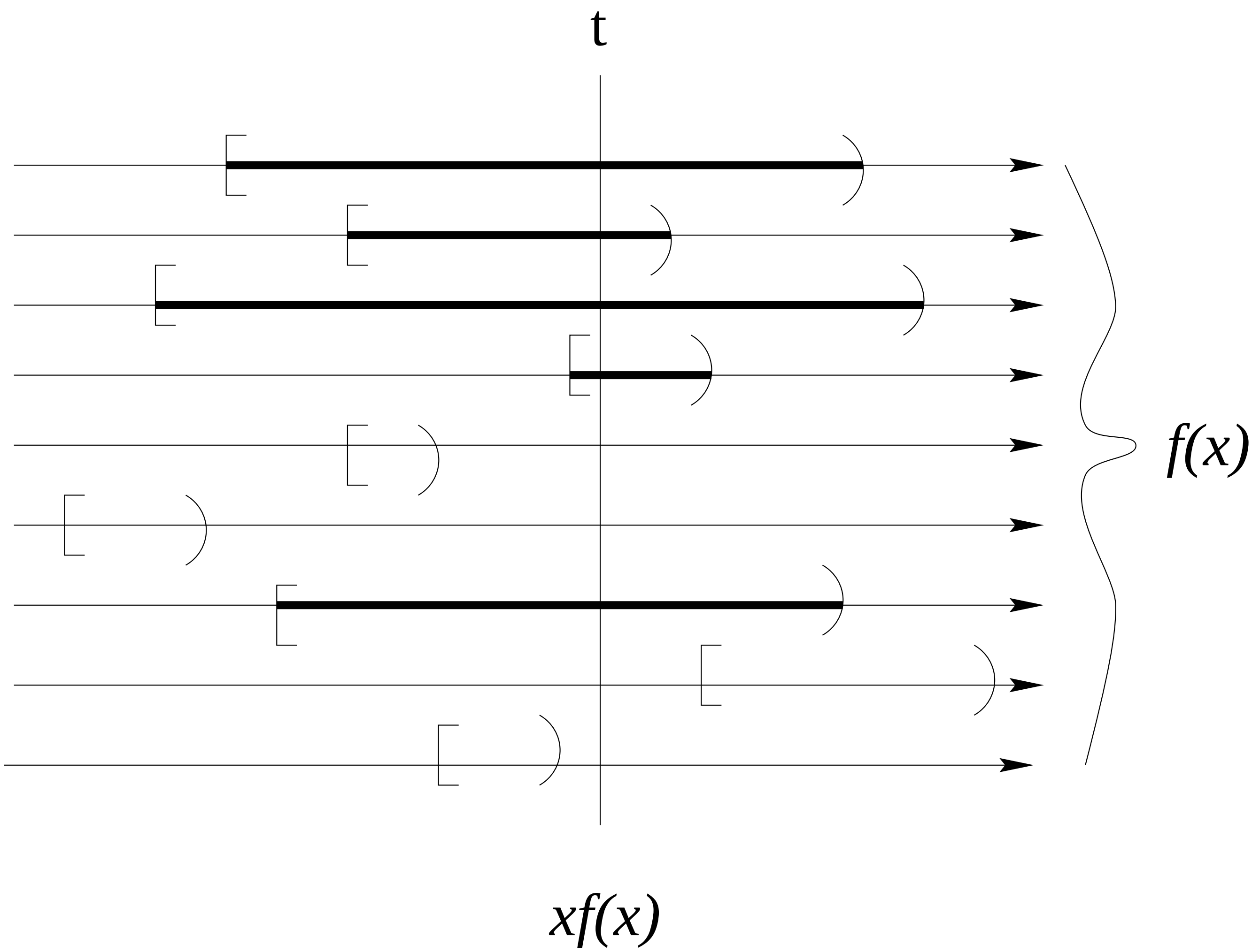
Mailing Address:
Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario, Canada.
E-mail:
cgsmall@uwaterloo.ca

Tilted densities.

- Let $f(x)$ be a probability density and $w(x)$ a nonnegative weight function. We define the *w-tilted* version of f to be the density

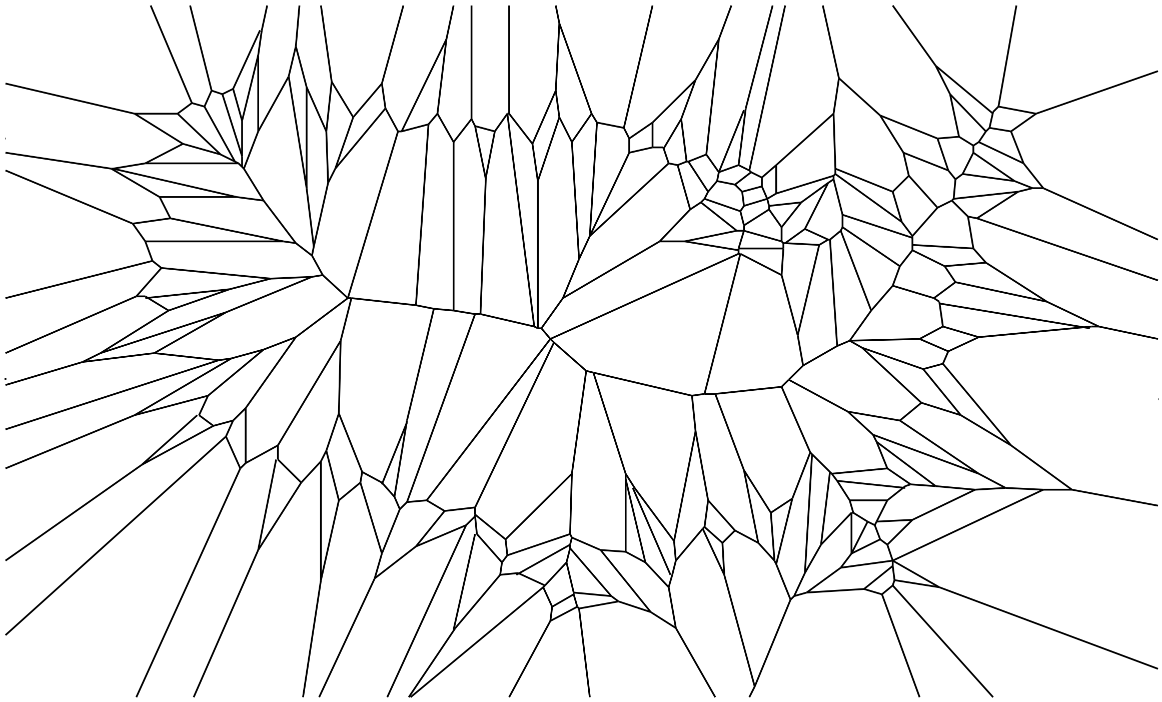
$$f_w(x) = \frac{w(x)\, f(x)}{\int w(x)\, f(x)\, dx} = \frac{w(x)\, f(x)}{E_f[w(X)]}$$

provided the denominator is finite.



Length-biased Sampling

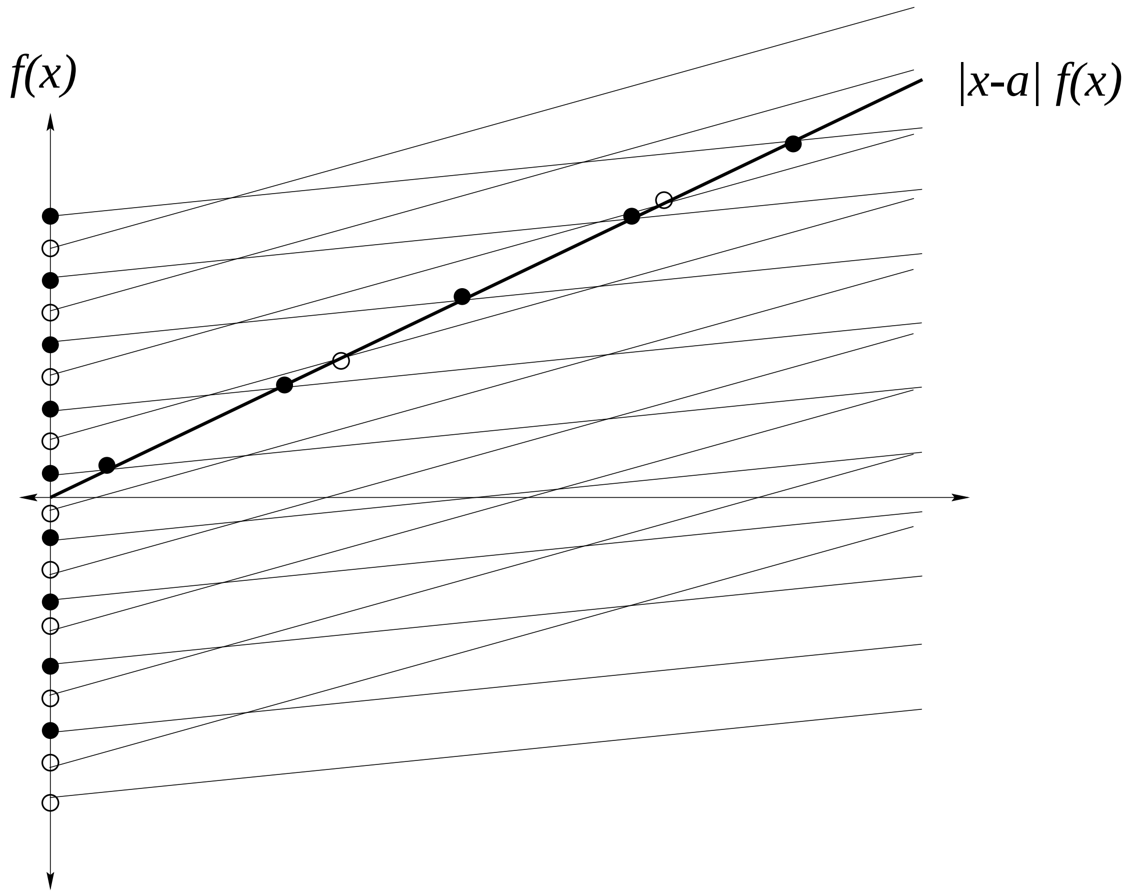
In length-biased sampling the weight function is $w(x) = x$.



Area-biased Sampling

- A polygon chosen at random from a population of polygons has area A
  - density $f(a)$.

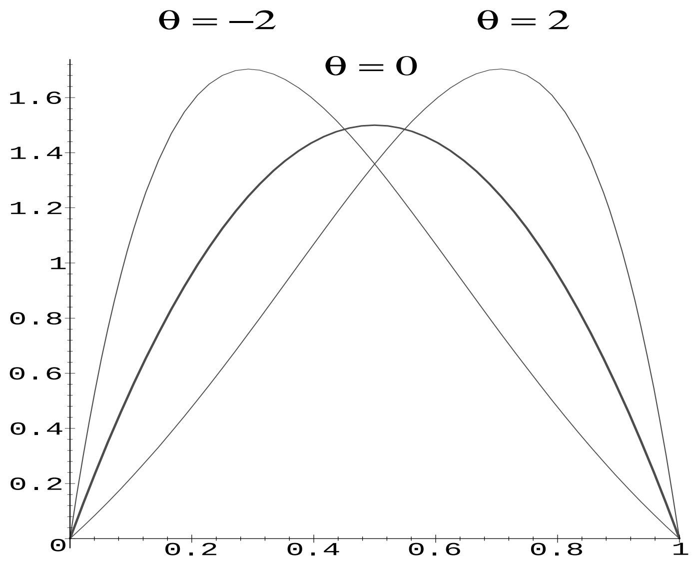
- Selecting a polygon by throwing a random dart at the region produces a random polygon with area A
  - density proportional to $a\,f(a)$.



Difference-Biased Sampling

- Estimate the distribution of velocities of cars driving along a road.
- We observe the cars on a fixed length of road at a given time:
  - density $f(x)$.
- Moving at velocity a along the road for a fixed time period we observe cars that we pass or which pass us:
  - density proportional to $|x - a|\, f(x)$.

Two exponential tilts of a Beta(2,2) density:



- The best known tilt is the *exponential tilt*,

$$f(x;\theta) = \exp[\theta\, x - K(\theta)]\, f(x)$$

  where $K(\theta)$ is the cumulant generating function of $f(x)$.

- The exponential tilt can be used for
  - The embedding of $f(x)$ in a linear exponential family;
  - The derivation of the saddlepoint approximation for $f(x)$ as a tilted Edgeworth expansion.

- It will be useful to regard a particular tilting operation as an *operator* on an appropriate space of density functions:

$$T_w \ : \ f \mapsto f_w$$

- Interpreting a tilt as an operator is consistent with the principle that a tilt is often the result of a transformation in the physical context of the model or a transformation in the sampling mechanism.

> By defining parameters with such operators, we interpret parameters *structurally*, separating them from the error distributions associated with "random noise."
>
> This takes us away from the Fisherian subordination of a parameter as the *index* of a family of distributions.

Tilted Expectations and Likelihood Tilts

- Suppose $w(x)$ is a tilting function that is standardised so that
$$E_f[w(X)] = 1.$$
The tilted expectation of a given function $h(x)$ is
$$E_{f_w}[h(X)] = E_f[w(X)\, h(X)].$$

- This is used for Monte Carlo studies when we wish to approximate the expectation $E_{f_w}[h(X)]$ but cannot generate random values from f_w. The density f is usually chosen so that simulation from f is feasible and the variance of $w(X)h(X)$ is not very high.

- For a parametric model with a family of densities $f(x;\theta)$, the tilting function is seen to be a *likelihood ratio*, and the tilting operation becomes a *likelihood tilt*. With parametric models it is natural to tilt functions $h(\theta, X)$ of both θ and X:
$$E_{\theta'}[h(\theta, X)] = E_\theta[\Lambda\, h(\theta, X)]$$
where Λ is the likelihood ratio
$$\Lambda = \frac{L(\theta')}{L(\theta)}.$$

- Note that Λ is automatically standardised because
$$E_\theta(\Lambda) = \int \frac{L(\theta')}{L(\theta)}\, L(\theta) = \int L(\theta') = 1.$$

- Also of interest is the *centred likelihood ratio*

$$\bar{\Lambda} = \frac{L(\theta')}{L(\theta)} - 1,$$

which is an unbiased estimating function in the sense that

$$E_\theta\left[\bar{\Lambda}\right] = 0.$$

- A limiting case of $\bar{\Lambda}$ occurs when $\theta' \to \theta$, and the centred likelihood ratio is renormalised:

$$\begin{aligned}\frac{1}{\theta' - \theta}\bar{\Lambda} &\to U(\theta, X) \\ &= \frac{\partial}{\partial \theta} \log L(\theta, X)\end{aligned}$$

which is the *score function.*

- Tilting an expectation with respect to the score yields

$$\begin{aligned}E_\theta\left[U(\theta, X)h(\theta, X)\right] &= \int h(\theta)\frac{\frac{\partial}{\partial\theta}L(\theta)}{L(\theta)}L(\theta) \\ &= \int h(\theta, x)\frac{\partial}{\partial\theta}L(\theta) \\ &= \frac{\partial}{\partial\theta}\int h(\theta, x)L(\theta) - \int \frac{\partial}{\partial\theta}h(\theta, x)\,L(\theta) \\ &= \frac{\partial}{\partial\theta}E_\theta[h(\theta, X)] - E_\theta\left[\frac{\partial}{\partial\theta}h(\theta, X)\right]\end{aligned}$$

- An important special case of this formula occurs when $E_\theta[h(\theta, X)]$ is functionally independent of θ. Then

$$\begin{aligned}E_\theta\left[U(\theta, X)h(\theta, X)\right] &= -E_\theta\left[\frac{\partial}{\partial\theta}h(\theta, X)\right] \\ &= \left\{\frac{\partial}{\partial\theta'}E_{\theta'}\left[h(\theta, X)\right]\right\}_{\theta'=\theta}\end{aligned}$$

Likelihood Functionals

- Associated with each likelihood ratio is a *likelihood functional*. Let $\mathcal{H}$ denote the class of all functions $h(\theta, X)$ such that

$$E_\theta[h(\theta, X)] \text{ is functionally independent of } \theta;$$

  and

$$E_\theta\left[h^2(\theta, X)\right] < \infty.$$

  Then $\mathcal{H}$ is a complete inner product space.

- Suppose also that $E_\theta[\Lambda^2] < \infty$. We can consider Λ to be a function of θ, and think of θ' as a subscript: $\Lambda = \Lambda_{\theta'}(\theta)$. Since $E_\theta[\Lambda] = 1$, the likelihood ratio Λ is an element of $\mathcal{H}$.

- Then we define the *likelihood functional*

$$\Lambda^* \;:\; h \quad\mapsto\quad E_{\theta'}[h(\theta, X)] = E_\theta\left[\Lambda\, h(\theta, X)\right].$$

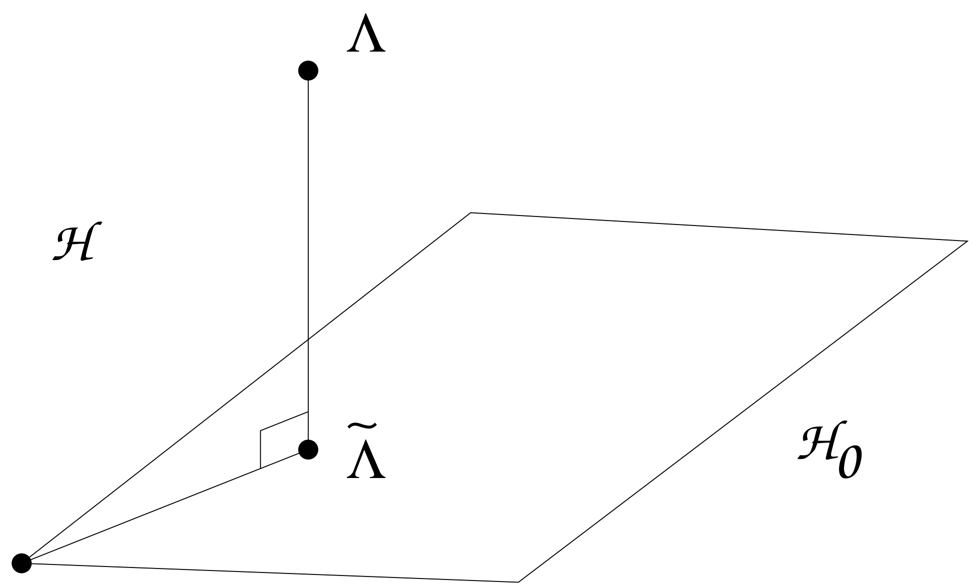
- Notes on this functional:

  - The mapping $\Lambda \mapsto \Lambda^*$ maps likelihood ratios in $\mathcal{H}$ to elements of its dual space $\mathcal{H}^*$.

  - Λ^* is a continuous (i.e., bounded) linear functional, with norm

$$||\Lambda^*|| = ||\Lambda|| = \sqrt{E_\theta[\Lambda^2]}.$$

The likelihood functionals have some advantages over likelihood ratios.
In particular, the likelihood functionals generalise to settings where likelihood ratios do not exist.



- Suppose $\mathcal{H}_0$ is a closed subspace of $\mathcal{H}$.
- Then the functional Λ^* will be defined on $\mathcal{H}_0$ even if Λ is not an element of $\mathcal{H}_0$.
- The norm of the restriction of Λ^* to $\mathcal{H}_0$ will no longer be $||\Lambda||$ but will be $||\widetilde{\Lambda}||$, where $\widetilde{\Lambda}$ is the projection of Λ into $\mathcal{H}_0$.
- Within $\mathcal{H}_0$, the function $\widetilde{\Lambda}$ therefore acts as an *analog* of a likelihood ratio, serving as a substitute for the latter in the *Riesz Representation Theorem*.

Application to Semiparametrics

- Let $X_j,\ j = 1, \ldots, n$ be independent, with means $\mu_j(\theta)$ and variances $\sigma_j^2(\theta)$ respectively.
- Suppose $\mathcal{H}_0$ consists of all functions of the form

$$h_{j_1 \cdots j_k}(\theta) = \prod_{i=1}^{k} [X_{j_i} - \mu_{j_i}(\theta)]$$

  where the product is taken over any subset of indices, together with all *linear combinations* of such functions and the function $h_0 \equiv 1$.
- Let $\Lambda = \Lambda_{\theta'}(\theta)$ be a likelihood ratio.
- The the projection of Λ into the space of multiples of $h_{j_1 \cdots j_k}$ is

$$\frac{\langle \Lambda, h_{j_1 \cdots j_k} \rangle}{||h_{j_1 \cdots j_k}||^2} h_{j_1 \cdots j_k}(\theta) = \prod_{i=1}^{k} \left[\frac{\mu_{j_i}(\theta') - \mu_{j_i}(\theta)}{\sigma_{j_i}^2(\theta)} \right] h_{j_1 \cdots j_k}(\theta).$$

- As the basis function $h_{j_1 \cdots j_k}$ are all orthogonal, the projection of Λ into the space $\mathcal{H}_0$ is found by adding up these components, yielding:

$$\widetilde{\Lambda} = 1 + \sum_{j_1, \cdots j_k} \left\{ \prod_{i=1}^{k} \left[\frac{\mu_{j_i}(\theta') - \mu_{j_i}(\theta)}{\sigma_{j_i}^2(\theta)} \right] h_{j_1 \cdots j_k}(\theta) \right\}.$$

  where the sum is over all subsets of indices.

- But this expression is simply the expansion of

$$\widetilde{\Lambda}_{\theta'}(\theta) = \prod_{j=1}^{n} \left\{ 1 + \frac{\mu_j(\theta') - \mu_j(\theta)}{\sigma_j^2(\theta)} [X_j - \mu_j(\theta)] \right\}.$$

  which is the Riesz representation of the likelihood functional in $\mathcal{H}_0$.

- Note that as $\theta' \to \theta$,

$$\frac{1}{\theta' - \theta} \widetilde{\Lambda}_{\theta'}(\theta) \ \to \ \sum_{j=1}^{n} \frac{\frac{\partial}{\partial\theta}\mu_j(\theta)}{\sigma_j^2(\theta)} [X_j - \mu_j(\theta)],$$

  which is the quasi-score.

- So the quasi-score is the Riesz representation of the score functional in $\mathcal{H}_0$.

- This *projected likelihood ratio* behaves like a likelihood ratio in some respects, except that it can go negative – typically when θ' is not in a $n^{-1/2}$-neighbourhood of θ.

- The function $\widetilde{\Lambda}$ does not directly define a projected likelihood for the parameter space since we cannot write

$$\widetilde{L}(\theta') = \widetilde{\Lambda}_{\theta'}(\theta) \, \widetilde{L}(\theta).$$

- However, if $\hat{\theta}$ is the unique quasi-MLE, then

$$\widetilde{L}(\theta) = \widetilde{\Lambda}_{\theta}(\hat{\theta})$$

  provides a consistent support function on the parameter space that is maximised at $\hat{\theta}$ where $\widetilde{L}(\hat{\theta}) = 1$.

- To illustrate this function, consider the IID model where $E(X) = \theta$ and $\mathrm{Var}(X) = 1$. Then $\hat{\theta} = \bar{X}$. So

$$\widetilde{L}(\theta) = \prod_{j=1}^{n} \left\{1 + (\theta - \bar{X})(X_j - \bar{X})\right\}.$$

Expanding we get

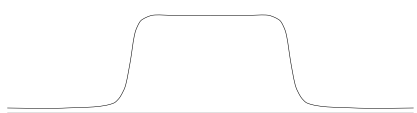
$$\begin{aligned}\widetilde{L}(\theta) &= 1 + (\theta - \bar{X}) \sum_{j=1}^{n} (X_j - \bar{X}) \\ &\qquad + (\theta - \bar{X})^2 \sum_{i \neq j} (X_i - \bar{X})(X_j - \bar{X}) + \cdots \\ &= 1 - (\theta - \bar{X})^2 (n-1) S^2 + \cdots\end{aligned}$$

- So a "likelihood" region for θ has the form $\widetilde{L}(\theta) \geq a$, or

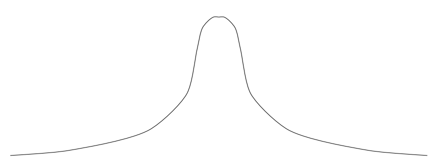
$$\bar{X} - \sqrt{\frac{1-a}{n-1}} S^{-1} \;\leq\; \theta \;\leq\; \bar{X} + \sqrt{\frac{1-a}{n-1}} S^{-1},$$

approximately.

- Note that as the sample variance increases, the interval becomes smaller. This is typical of platykurtic location models. (We are not assuming normality here.)



Platykurtic density



Leptokurtic density

- When will $\widetilde{\Lambda}$ be a true likelihood ratio ?
- For this, we must have

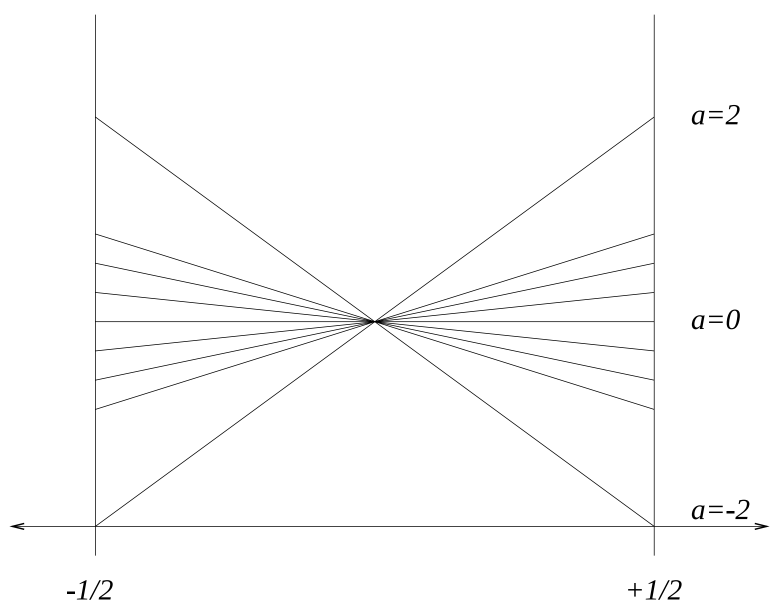
$$f(x;\theta') = f(x;\theta)\left\{1 + \frac{\mu' - \mu}{\sigma^2}[x - \mu]\right\}$$

- If there is a member $f(x)$ of this family such that $\int x f(x) dx = 0$, then we can write

$$f(x;\theta) = f(x) + a(\theta)\, x f(x)$$

where we must have $1 + x\, a(\theta) \geq 0$ on the support of f.

- So $f(x;\theta)$ is a *mixture* of a density f and a signed-length biased version of f.
- Example:



Likelihood Functionals using Edgeworth Expansions

- It is also possible to express likelihood functionals using generalised Edgeworth expansions.

- Let $f(x;\theta)$ be a family of densities, with cumulant generating function $K(t;\theta)$, respectively. We write

$$K(t;\theta') - K(t;\theta) = \sum_{j=1}^{\infty} \frac{\kappa_j(\theta') - \kappa_j(\theta)}{j!} t^j \; .$$

So the MGF of $f(x;\theta')$ is

$$\begin{aligned} M(t;\theta') &= \exp\left[K(t;\theta') - K(t;\theta)\right] M(t;\theta) \\ &= \exp\left[\sum_{j=1}^{\infty} \frac{\kappa_j(\theta') - \kappa_j(\theta)}{j!} t^j\right] M(t;\theta) \end{aligned}$$

- Doing a Taylor expansion on the exponential (Kendall & Stuart, McCullagh 1987, Kolassa 1994), we have

$$\exp\left[\sum_{j=1}^{\infty} \frac{\kappa_j(\theta') - \kappa_j(\theta)}{j!} t^j\right] = \sum_{k=0}^{\infty} \frac{\mu_j^*(\theta',\theta)}{j!} t^j \; ,$$

where the $\mu_j^*(\theta',\theta)$ are called pseudo-moments. So

$$M(t;\theta') = \sum_{k=0}^{\infty} \frac{\mu_j^*(\theta',\theta)}{j!} \left\{t^j \, M(t;\theta)\right\}$$

- Taking the inverse Laplace transform of this equation, and using the fact that $t^j\, M(t;\theta)$ inverts to $(-d/dx)^j\, f(x;\theta)$, we get

$$f(x;\theta') = \sum_{k=0}^{\infty} \frac{\mu_j^*(\theta',\theta)}{j!} \left\{ \left(-\frac{d}{dx}\right)^j f(x;\theta) \right\} \quad .$$

- Wrapping the exponential back up, we have

$$f(x;\theta') = \exp\left[\sum_{j=1}^{\infty} \frac{\kappa_j(\theta') - \kappa_j(\theta)}{j!} \left(-\frac{d}{dx}\right)^j \right] f(x,\theta) \quad .$$

- So we can formally represent the likelihood functional as

$$\Lambda^* = \exp\left[\sum_{j=1}^{\infty} \frac{\kappa_j(\theta') - \kappa_j(\theta)}{j!} \left(-\frac{d}{dx}\right)^j \right] \quad .$$

- This formula allows us to write likelihood functionals in terms of the cumulants of the model – in the spirit of semiparametrics – without explicitly using density functions.

- To illustrate this formula, consider $X \sim N(0,\theta)$. Then $\kappa_2(\theta) = \theta$ and $\kappa_j(\theta) = 0$ for all $j \neq 2$. So

$$\Lambda^* = \exp\left(\frac{\theta' - \theta}{2} \frac{d^2}{dx^2} \right) \quad .$$

In the limit as $\theta' \to \theta$, the score functional is found to be

$$\frac{1}{2} \frac{d^2}{dx^2}$$

which is easily recognised as the Laplacian operator for the variation coefficient of the normal distribution – i.e., the diffusion coefficient in the heat equation, etc.