5.7 Gram-Charlier

A normalized form for the assumed distribution of the data x is

$$f(x,\sigma,\alpha) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}} \left(1 + \alpha(1 - \left(\frac{x}{2}\right)^2)\right)$$

where σ is the standard deviation of the basic Gaussian and α measures the contribution of the Gram-Charlier term.
For a set of N data X_i the likelihood function is a product, so that

$$\text{prob}(\alpha,\sigma|X_i) \propto \prod_i f(X_i,\sigma,\alpha).$$

Because f is a sum, the product gets very complicated. To investigate further, assume $\alpha << 1$ and keep only terms linear in the product. This gives

$$\text{prob}(\alpha,\sigma|X_i) \propto \frac{1}{(\sqrt{2\pi}\sigma)^N} \left(e^{\frac{-\sum_i X_i^2}{2\sigma^2}} + \alpha \sum_k (1 - X_k^2) e^{\frac{-\sum_{i\neq k} X_i^2}{2\sigma^2}} \right) \text{prob}(\alpha)\text{prob}(\sigma)$$

Marginalizing out α can be more or less complicated to taste – the prior is open to debate. Taking a uniform distribution between 0 and t for α, so $\text{prob}(\alpha) = 1/t$, the relevant factors are simple, giving

$$\text{prob}(\sigma|X_i) \propto \left(e^{\frac{-\sum_i X_i^2}{2\sigma^2}} + \frac{t}{2} \sum_k (1 - X_k^2) e^{\frac{-\sum_{i\neq k} X_i^2}{2\sigma^2}} \right) \text{prob}(\sigma).$$

The Jeffreys prior is appropriate for $\text{prob}(\sigma)$. Graphing the distribution for small sets of Gaussian data, taking $t \simeq 0.1$, we see a definite tendency for smaller values of the most likely σ than we get for $\alpha = 0$. The extra term in the Gram-Charlier expansion has a tendency to absorb some of the spread in the data.
The odds on including the Gram-Charlier term are given by the ratio

$$\mathcal{O} = \frac{\int \text{prob}(\sigma|X_i, \text{finite}\,\alpha)\, d\sigma}{\int \text{prob}(\sigma|X_i, \alpha = 0)\, d\sigma}$$

which is the ratio of the weights of evidence. These integrations can be done numerically. For smallish sets of data ($N \simeq 30$) we find typical odds of about 10 to 1 (assuming equal prior probabilities for the two hypotheses) in favour of including the term. While the effect on the distribution of small values of α may seem limited, it can affect significances in the tails by a per cent. or so.