Теорема.
Пусть задана выборка $\mathbf{X}, \mathbf{y}$ мощности m.
Пусть задана модель $\mathbf{f}(\mathbf{w}, \mathbf{X})$ и распределение q, апппроксимирующее апостериорное распределеине параметров $\mathbf{w}$ этой модели.
Рассмотрим выражение $\frac{1}{m}\text{ELBO}_\gamma$:

$$\frac{1}{m}\text{ELBO}_\gamma(\mathbf{X}, \mathbf{y}, q) = \frac{1}{m}\mathsf{E}_q \text{log} p(\mathbf{y}|\mathbf{X}, \mathbf{w}) - \frac{\gamma}{m}\text{KL}(q|p(\mathbf{w})),$$

где $\gamma > 0$.

Пусть $\frac{1}{m}\text{ELBO}_\gamma$ сходится п.н. при $m \to \infty$ к функции $L(q)$
(вообще, она еще от гиперпараметров зависит, но здесь это будет лишним, прим. Олег).

Тогда функция $\frac{1}{m_0}\text{ELBO}_1$ для выборки мощности $m_0 = \frac{m}{\gamma}$ из той же генеральной совокупности сходится почти наверно к этой же функции $L(q)$:

$$\frac{1}{m_0}\text{ELBO}_1(\hat{\mathbf{X}}, \hat{\mathbf{y}}, q) \to^{\text{п.н.}} L(q),$$

где $|\hat{\mathbf{X}}| = m_0$.

Доказательство. Рассмотрим величину $\frac{1}{m}\text{ELBO}_\gamma$:

$$\frac{1}{m}\text{ELBO}_\gamma(\mathbf{X}, \mathbf{y}) = \frac{1}{m}\mathsf{E}_q \text{log} p(\mathbf{y}|\mathbf{X}, \mathbf{w}) - \frac{\gamma}{m}\text{KL}(q|p(\mathbf{w})).$$

По УЗБЧ:

$$\frac{1}{m}\text{ELBO}_\gamma(\mathbf{X}, \mathbf{y}) \to^{\text{п.н.}}_{m\to\infty} \mathsf{E}_{\mathbf{X}}\mathsf{E}_q \text{log} p(\mathbf{y}|\mathbf{X}, \mathbf{w}) - \frac{\gamma}{m}\text{KL}(q|p(\mathbf{w})) = L(q).$$

Аналогично рассмотрим $\frac{1}{m_0}\text{ELBO}_1$ для выборки мощностью $m_0 = \frac{m}{\gamma}$:

$$\frac{1}{m_0}\text{ELBO}_1(\hat{\mathbf{X}}, \hat{\mathbf{y}}) \to^{\text{п.н.}}_{m\to\infty} \mathsf{E}_{\mathbf{X}}\mathsf{E}_q \text{log} p(\mathbf{y}|\mathbf{X}, \mathbf{w}) - \frac{1}{m_0}\text{KL}(q|p(\mathbf{w})) =$$

$$= \mathsf{E}_{\mathbf{X}}\mathsf{E}_q \text{log} p(\mathbf{y}|\mathbf{X}, \mathbf{w}) - \frac{\gamma}{m}\text{KL}(q|p(\mathbf{w})) = L(q),$$

предельные функции совпадают, что и требовалось доказать.

Интерпретация: для достаточно большого m и $\gamma > 0, \gamma \neq 1$ оптимизация параметров и гиперпараметров эквивалентна оптимизации ELBO для выборки другой мощности:

$$\max_q \text{ELBO}_\gamma(\mathbf{X}, \mathbf{y}, q) \propto \max_q \frac{1}{m}\text{ELBO}_\gamma(\mathbf{X}, \mathbf{y}, q) \sim \max_q \frac{1}{m_0}\text{ELBO}_1(\hat{\mathbf{X}}, \hat{\mathbf{y}}, q) \sim$$

$$\sim \max_q \text{ELBO}_1(\hat{\mathbf{X}}, \hat{\mathbf{y}}, q)$$

К примеру, оптимизация ELBO_γ при $\gamma > 1$ эквивалентна оптимизации ELBO для выборки меньшей мощности (и бОльшего вклада априорного распределения в оптимизацию).