Model selection for the gestalt model

1 Approximating the likelihood

In order to evaluate the log-likelihood for cross validation steps, we have to use an approximation.

$$
\begin{aligned}
p(X \mid C_{1..k}) = \prod_{n=1}^{N} \iiint_{-\infty}^{\infty} p(x_n \mid v_n, z_n) p(v_n \mid g_n) p(g_n) p(z_n) \mathrm{d}v_n \mathrm{d}g_n \mathrm{d}z_n = \\
= \prod_{n=1}^{N} \iint_{-\infty}^{\infty} p(g_n) p(z_n) \int_{-\infty}^{\infty} p(x_n \mid v_n, z_n) p(v_n \mid g_n) \mathrm{d}v_n \mathrm{d}g_n \mathrm{d}z_n
\end{aligned}
\tag{1}
$$

where the double integral may be approximated by averaging over samples from $p(g)$ and $p(z)$

$$
p(X \mid C_{1..k}) \approx \prod_{n=1}^{N} \frac{1}{L_1 L_2} \sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \int_{-\infty}^{\infty} p(x_n \mid v, z^{l_2}) p(v \mid g^{l_1}) \mathrm{d}v \tag{2}
$$

and as samples of g and z are independent from each other we can set $L = L_1 L_2$, and thus

$$
p(X \mid C_{1..k}) \approx \prod_{n=1}^{N} \frac{1}{L} \sum_{l=1}^{L} \int_{-\infty}^{\infty} p(x_n \mid v, z^{l}) p(v \mid g^{l}) \mathrm{d}v \tag{3}
$$