Mass Spectrometer

- Mass spectrometry is an analytical technique used to measure the mass-to-charge ratio (M/Z) of ionized proteins.
- Example: MALDI and SELDI.

Scientific Interests

Peaks in spectrum correspond to proteins.
- Feature Selection
  - Peak presence/absence, different abundance.
  - Classification, prediction.

Data Issues

- Calibration
  - Convert time of flight to mass charge. Align multiple mass spectra.
- Baseline
  - Cause: Matrix molecules; detector overload.
  - So: Build into the model, or subtract it.
- Normalization
  - Ciphergen automatically rescales the spectrum. Model shouldn’t be sensitive to vertical scaling.

Noise Model

- As a function of mean intensity?
- As a function of time?

Resolution

- If \( \Delta \tau \) is the smallest time difference for which two peaks with time \( \tau \) and \( \tau + \Delta \tau \) are resolved, then the resolution is
  \[
  \rho = \frac{\tau}{\Delta \tau}.
  \]
- \( \Delta \tau \) is full width at half-maximum (FWHM).
  If \( t \sim N(\tau, \sigma^2) \), then \( \rho = \frac{\tau}{2\sigma\sqrt{\log 2}} \).

Model Specification

\[
Y_t = \sum_{j=1}^{J} \beta_j N_{\text{true}}(t; \tau_j, \sigma^2(\tau_j, \rho)) + \epsilon_t,
\]
\[
\epsilon_t \sim N(0, \phi^{-1}a^2(\tau^2 + \sigma^2)).
\]

Priors:

- Expected time of flight: \( \tau_j \sim U[T_{\text{min}}, T_{\text{max}}] \).
- Abundance: \( \beta_j \sim N(0, \nu_0 \sigma^2_0) \).
- Number of peaks: \( J \sim N_{\text{eg}}(r_0, p_0) \).
- Resolution: \( \rho \sim N_{\text{isc}}(p_0, V_0) \).
- Noise: \( \phi \sim \text{Ga}(\nu_0/2, \nu_0 s^2_0/2) \), \( a \sim N(\alpha_0, V_0) \).

Calculate Marginal Likelihood

Model:

\[
Y = X\beta + \epsilon , \quad \epsilon \sim N(0, \phi^{-1}W^{-1}_a).
\]

Features

- Interpretable parameters
- Time varying noise model.
- Compact kernel leads to efficient updating.
- Informal proposal for RJ MCMC.

Future Directions

- Model for baseline.
- Robust priors on the coefficients \( \beta \).
- Modeling dependence: hierarchical structure
  - Multiple charged peaks, isotopes.
  - Clusters: shared features.
  - Longitudinal trends.