Two-Dimensional Problems

Consider a model with two parameters: \( x, y \)

The posterior probability is:

\[ P(x, y | D, I) \]

The solution \( x_0, y_0 \) will be given by the solution to the two simultaneous equations

\[
\frac{\partial P}{\partial x} \bigg|_{x_0, y_0} = 0 \quad \frac{\partial P}{\partial y} \bigg|_{x_0, y_0} = 0
\]

or equivalently

\[
\frac{d \log P}{dx} \bigg|_{x_0, y_0} = 0 \quad \frac{d \log P}{dy} \bigg|_{x_0, y_0} = 0
\]

This will give two equations

(a) \( f_x(x_0, y_0) = 0 \)
(b) \( f_y(x_0, y_0) = 0 \)

Strategies

1. Solve for \( x_0 \) and \( y_0 \) analytically \( \Rightarrow \) ANALYTIC SOLN

2. Solve for \( x_0 \) and \( y_0 \) by iterating \( \Rightarrow \) ITERATIVE FIXED POINT SOLN
   
   Guess \( x_0 \), solve (a) for \( y_0 \), solve (b) for \( x_0 \) and repeat
   
   DOESN'T ALWAYS WORK

3. Solve for \( x_0 \) and \( y_0 \) numerically
Uncertainty in Two-D Problems

Again, we take a Taylor series expansion

\[ L = \log P(x, y | I, \theta) \]

\[ L = L(x_0, y_0) + \frac{\partial L}{\partial x} \bigg|_{x_0, y_0} (x - x_0) + \frac{\partial L}{\partial y} \bigg|_{x_0, y_0} (y - y_0) + \]

\[ + \frac{1}{2} \left[ \frac{\partial^2 L}{\partial x^2} \bigg|_{x_0, y_0} (x - x_0)^2 + \frac{\partial^2 L}{\partial y^2} \bigg|_{x_0, y_0} (y - y_0)^2 \right] + \ldots \]

Recall that \( \frac{\partial^2 L}{\partial x \partial y} = \frac{\partial^2 L}{\partial y \partial x} \)

We can write the quadratic part in matrix notation

\[ Q = \begin{bmatrix} x - x_0 & y - y_0 \end{bmatrix} \begin{bmatrix} A & C \\ C & B \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \]

where

\[ A = \frac{\partial^2 L}{\partial x^2} \bigg|_{x_0, y_0}, \quad B = \frac{\partial^2 L}{\partial y^2} \bigg|_{x_0, y_0}, \quad C = \frac{\partial^2 L}{\partial x \partial y} \bigg|_{x_0, y_0} \]

Taking the Exponential

\[ P(x, y | I, \theta) \propto \exp \left[ -\frac{1}{2} (x - x_0, y - y_0)^T \begin{bmatrix} A & C \\ C & B \end{bmatrix} (x - x_0, y - y_0) \right] \]

\[ \exp \left[ -\frac{Q}{2} \right] \]

\text{Hessian}

\text{Matrix of 2nd Partial Derivatives}
Uncertainty in Two-D Problems

To find the uncertainty in our estimate of $x_0$, we first marginalize out $y$. This will give us a posterior probability $x$.

$$P(x|D,i) = \int P(x,y|D,i) \, dy$$

$$= \int_{-\infty}^{+\infty} K \exp \left[ \frac{1}{2} (A(x-x_0)^2 + B(y-y_0)^2 + 2C(x-x_0)(y-y_0)) \right] \, dy$$

Let $x' = x - x_0 \quad dx' = dx$
$y' = y - y_0 \quad dy' = dy$

$$= K \int_{-\infty}^{+\infty} \exp \left[ \frac{1}{2} (A x'^2 + B y'^2 + 2C x' y') \right] \, dy'$$

$$= K \exp \left[ -\frac{C^2}{B} x^2 \right] \int_{-\infty}^{+\infty} \exp \left[ -\frac{C^2}{2B} x'^2 \right] \exp \left[ -\frac{1}{2} B (y' + \frac{C x'}{B})^2 \right] \, dy'$$

Complete the square

$$B y'^2 + 2C x' y' = B \left( y' + \frac{C x'}{B} \right)^2 - \frac{C^2}{B} x'^2$$

$$= K \exp \left[ -\frac{C^2}{2B} x^2 \right] \exp \left[ -\frac{1}{2} B (y' + \frac{C x'}{B})^2 \right] \, dy'$$
Uncertainty in Two-D Problems

\[ P(x, t, y, s) = K \exp \left( \frac{(A - \frac{C^2}{2B})x^2}{2} \right) \sqrt{\frac{1}{\pi}} \int_{-\infty}^{\infty} \exp \left( \frac{1}{2B} \left( y + \frac{C}{8} \right)^2 \right) dy \]

Let \( u = y + \frac{C}{8} \)

\[ = K \exp \left( \frac{(A - \frac{C^2}{2B})x^2}{2} \right) \sqrt{\frac{1}{\pi}} \int_{-\infty}^{\infty} \exp \left( \frac{1}{2Bu^2} \right) du \]

Let \( \sigma^2 = \frac{1}{B} \) then

\[ = K \exp \left( \frac{(A - \frac{C^2}{2B})x^2}{2} \right) \sqrt{\frac{1}{\pi}} \int_{-\infty}^{\infty} \exp \left( -\frac{u}{2\sigma^2} \right) du \]

\[ = K \exp \left[ \frac{1}{2} \left( \frac{AB - C^2}{B} \right) x^2 \right] \]

\[ \propto \exp \left[ -\frac{1}{2} \left( \frac{AB - C^2}{-B} \right) \left( X - X_o \right)^2 \right] \]

\[ \Rightarrow \sigma_X = \sqrt{\frac{-B}{AB - C^2}} \]

Similarly,

\[ \sigma_Y = \sqrt{\frac{-A}{AB - C^2}} \]
Uncertainty in Two-D Problems

Consider the Variance of $x$

$$\text{Var} \ x = \langle (x-x_0)^2 \rangle = \iint (x-x_0)^2 P(x,y|D,I) \, dx \, dy$$

we already did the integral over $y$

$$= \int (x-x_0) K' \text{Exp} \left[ -\frac{1}{2\sigma_x^2} (x-x_0)^2 \right]$$

$$\text{Var} \ x = \sigma_x^2$$

The covariance of $x$ and $y$ describes how the parameters $x$ and $y$ are correlated.

$$\sigma_{xy} = \langle (x-x_0)(y-y_0) \rangle$$

$$= \iint (x-x_0)(y-y_0) P(x,y|D,I) \, dx \, dy$$

For our 2D Gaussian, this is

$$= \frac{C}{AB-C^2}$$

**Covariance Matrix**

$$\text{COV} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \frac{1}{AB-C^2} \begin{pmatrix} -B & C \\ C & -A \end{pmatrix} = -\left( \begin{pmatrix} A & C \\ C & B \end{pmatrix} \right)^{-1}$$

Determinant of

$$\begin{pmatrix} A & C \\ C & B \end{pmatrix}$$
Covariance in 2D

Since

\[
\text{COV} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = - \begin{pmatrix} A & C \\ C & B \end{pmatrix}^{-1} = \frac{1}{AB - C^2} \begin{pmatrix} -B & C \\ C & -A \end{pmatrix}
\]

A catastrophe occurs when \( C^2 = AB \), \( C = \pm \sqrt{AB} \)

- The determinant is zero
- The matrix is singular

The ellipse becomes infinitely thin and infinitely long oriented at an angle \( \pm \tan^{-1} \sqrt{\frac{A}{B}} \) with the x-axis.

In this case we can only know a linear combination of \( x \) and \( y \). They cannot be disentangled.

Only a prior probability can rectify this situation or new relevant data.
Covariance in 2-D

$$\text{COV} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = -\begin{pmatrix} A & C \\ C & B \end{pmatrix}^{-1} = -(\nabla \nabla \psi)^{-1}$$

$$= -\begin{pmatrix} \frac{\partial^2 \text{Log } P}{\partial x^2} \bigg|_{x_0, y_0} & \frac{\partial^2 \text{Log } P}{\partial x \partial y} \bigg|_{x_0, y_0} \\ \frac{\partial^2 \text{Log } P}{\partial y \partial x} \bigg|_{x_0, y_0} & \frac{\partial^2 \text{Log } P}{\partial y^2} \bigg|_{x_0, y_0} \end{pmatrix}^{-1}$$

$$= -H^{-1}$$

Looking more closely at our quadratic approx...

The contours of the probability close to \((x_0, y_0)\) are ellipses.

For \((x_0, y_0)\) to be a maximum, \(\lambda_1 > 0\), \(\lambda_2 < 0\)

\[ \Rightarrow A < 0, \ B < 0 \text{ and } AB > C^2 \]

When \(C \neq 0\), the ellipse is skewed.
Covariance in 2D

When \( C > 0 \), the probability density is skewed. The estimates of \( x_0 \) and \( y_0 \) are not independent since:

\[
C = \frac{\partial^2 \text{Log } p}{\partial x \partial y}
\]

For this reason, we can't just take

\[
\sigma_{xx} = -\left( \frac{\partial^2 \text{Log } p}{\partial x^2} \right)^{-1}
\]

Instead, we must invert the entire matrix, which we found to be equivalent to marginalizing over \( y \) and then inverting the second derivative of the log marginal probability.

\[
\text{COV} = \begin{pmatrix}
\sigma_{xx} & \sigma_{xy} \\
\sigma_{yx} & \sigma_{yy}
\end{pmatrix} = -H^{-1}
\]

THREE CASES

- **Uncorrelated**
  - \( C = 0 \)
  - Better inference about \( x \), know \( y \)
  - \( \text{Corr } x, y \)

- **Positively Correlated**
  - \( y + mx = \text{Const} \)
  - Better inference \( y - mx \)
  - \( y - x \)

- **Negatively Correlated**
  - \( y + mx = \text{Const} \)
  - Better inference \( y + mx \)
  - \( y + x \)
Approximating the Hessian: 1D case

If it is too difficult, or impossible, to analytically compute the Hessian matrix,

One can easily generate a numeric approximation.

Definition of the Derivative

\[ f'(x_i) = \lim_{h \to 0} \frac{f(x_i + h) - f(x_i)}{h} \]

Approximation

\[ f'(x_i) \approx \frac{f(x_i + h) - f(x_i)}{h} \]

Perform a Taylor's Series approx of \( f(x_i + h) \)

\[ f(x_i + h) \approx f(x_i) + h f'(x_i) + \frac{h^2}{2} f''(x_i) + \mathcal{O}(h^3) \]

Also

\[ f(x_i - h) \approx f(x_i) - h f'(x_i) + \frac{h^2}{2} f''(x_i) + \mathcal{O}(h^3) \]

Solve for \( f'(x_i) \)

\[ f(x_i + h) - f(x_i - h) \approx 2h f'(x_i) \]

\[ \Rightarrow f'(x_i) = \frac{f(x_i + h) - f(x_i - h)}{2h} \]

Expanding out further we can find

\[ f''(x_i) = \frac{f(x_i + h) - 2f(x_i) + f(x_i - h)}{h^2} + \mathcal{O}(h^2) \]
Approximating the Hessian: 2D

\[ \frac{\partial^2 L}{\partial x^2} \bigg|_{x_0,y_0} = \frac{L(x_0+h, y_0) - 2L(x_0, y_0) + L(x_0-h, y_0)}{h^2} \]

Similarly,

\[ \frac{\partial^2 L}{\partial y^2} \bigg|_{x_0,y_0} = \frac{L(x_0, y_0+h) - 2L(x_0, y_0) + L(x_0, y_0-h)}{h^2} \]

\[ \frac{\partial^2 L}{\partial x \partial y} \bigg|_{x_0,y_0} = \frac{L(x_0+h, y_0+h) - L(x_0-h, y_0+h) - L(x_0+h, y_0-h) + L(x_0-h, y_0-h)}{4h^2} + O(h) \]

How big should \( h \) be?

The computer truncation error is about \( O(h^3) \)

where \( h \) is the smallest machine number

(as long as \( L \) is not too complicated)

For one, \( \frac{O(h)}{h^2} \approx O(h) \) ~ Formula truncation error

For two of them, \( \left( \frac{O(h)}{h^2} \right)^2 = h^3 \Rightarrow h \approx O(h^\frac{3}{2}) \)

In general for:

\[ \frac{\partial^2 L}{\partial x_i^2} \bigg|_{x_0} \] vary \( x_i \), only keep all other \( x_j \) (\( j \neq i \)) constant at \( x_0 \)

\[ \frac{\partial^2 L}{\partial x_i \partial x_j} \bigg|_{x_0, x_0} \] vary \( x_i \) and \( x_j \), only as in eqn above for \( x + y \).
Multi-Dimensional Generalization

We now label our parameters as \( X_i = \xi, x_2, \ldots, x_M \)

I will now deviate from Sierra and Skilling and use \( \hat{\cdot} \) to indicate the mode.

The optimal parameter vector

\[
\frac{\partial L}{\partial X_i} \bigg|_{\hat{X}_0} = 0 \quad \text{where } i = 1, 2, \ldots, M
\]

to give \( M \) simultaneous equations.

The Taylor series expansion is

\[
L = L(\hat{X}_0) + \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{\partial^2 L}{\partial X_i \partial X_j} \bigg|_{\hat{X}_0} (X_i - \hat{X}_i)(X_j - \hat{X}_j) + \ldots
\]

\( \Rightarrow \)

\[
P(\tilde{X} \mid D, I) = \frac{1}{Z} \exp \left[ \frac{1}{2} ( \tilde{X} - \tilde{X}_0 )^T \tilde{\Sigma} \Omega L(\tilde{X}) ( \tilde{X} - \tilde{X}_0 ) \right]
\]

where

\[
Z = (2\pi)^{M/2} \sqrt{|H|}
\]

Covariance matrix

\[
[\Sigma^2]_{ij} = \langle (X_i - \hat{X}_i)(X_j - \hat{X}_j) \rangle = \left[ (\tilde{\Omega} \tilde{\Sigma} \Omega L) \right]_{ij} = \left[ H^{-1} \right]_{ij}
\]

\( \Sigma^2_{xx} \) is the diagonal \((x, x)\) element of the inverse \(O \) of \(-H\)

is NOT the inverse of the diagonal element \((x, x)\) of \(-H\).