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 SOLUTION

2. Solve for \( x_0 \) and \( y_0 \) by iterating \( \Rightarrow \) ITERATIVE FIXED POINT SOLUTION
   - 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

\[ L = \log P(x, y | d, i) \]

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

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

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

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

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

Taking the Exponential

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

\[ \exp \left[ -\frac{1}{2} Q \right] \]

\[ \text{Hessian} \]
\[ \text{Matrix of 2nd Partial Derivatives} \]
Uncertainty in Two-Dimensional Problems

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

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

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

Let \( x' = x - x_0 \) \( dx' = dx \)
\( y' = y - y_0 \) \( dy' = dy \)

\[
= K \int \exp \left[ \frac{1}{2} (Ax'^2 + By'^2 + 2C x'y') \right] \, dy'
\]

\[
= K \exp \left[ \frac{1}{2} Ax'^2 \right] \int \exp \left[ \frac{1}{2} By'^2 \right] \, dy'
\]

Complete the square

\[
By'^2 + 2C x'y' = B (y' + \frac{C x'}{B})^2 - \frac{C^2}{B} x'^2
\]

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

\[ P(x, \theta, \tau) = K \exp \left[ \frac{(A \cdot \tau - c \cdot \theta^2)}{2} \right] \int_{-\infty}^{\infty} \exp \left[ \frac{1}{2} B (Y' + \frac{c'X'}{\theta})^2 \right] dY' \]

Let \( u = Y' + \frac{c'X'}{\theta} \)

\[ = K \exp \left[ \frac{(A \cdot \tau - c \cdot \theta^2)}{2} \right] \int_{-\infty}^{\infty} \exp \left[ \frac{1}{2} B u^2 \right] du \]

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

\[ = K \exp \left[ \frac{(A \cdot \tau - c \cdot \theta^2)}{2} \right] \int_{-\infty}^{\infty} \exp \left[ -\frac{u^2}{2 \sigma_x^2} \right] du \]

\[ = K' \exp \left[ \frac{1}{2} \left( \frac{AB - c^2}{\theta} \right) x'^2 \right] \]

\[ \propto \exp \left[ -\frac{1}{2} \left( \frac{AB - c^2}{-B} \right) (X - x_0)^2 \right] \]

\[ \implies \sigma_x = \sqrt{\frac{-B}{AB - c^2}} \]

Similarly

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

Consider the Variance of $X$

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

We already did the integral over $Y$

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

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

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

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

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

For our 2D Gaussian, this is

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

Covariance Matrix

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

\text{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 \( \pm \tan^{-1} \frac{A}{B} \) 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

\[
\text{Cov} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} = -(A \quad C)^{-1} = -(\nabla \nabla L)^{-1}
\]

\[
= -\begin{pmatrix} \frac{\partial^2 \log P}{\partial x^2} \bigg|_{x_0, y_0} & \frac{\partial^2 \log P}{\partial x \partial y} \bigg|_{x_0, y_0} \\ \frac{\partial^2 \log P}{\partial y \partial x} \bigg|_{x_0, y_0} & \frac{\partial^2 \log P}{\partial y^2} \bigg|_{x_0, y_0} \end{pmatrix}^{-1}
\]

\[
= -H^{-1}
\]

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, \(\lambda_1 < 0 \quad \lambda_2 < 0\)

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

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

When $C > 0$, the probability density $p$ is skewed. The estimates of $x_0$ and $y_0$ are not independent since

$$C = \frac{\partial^2 \log p}{\partial x \partial y}$$

For this reason, we can't just take

$$\Sigma_x^2 = -\left(\frac{\partial^2 \log p}{\partial x^2}\right)^{-1}$$

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.

$$\text{COV} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = -H^{-1}$$

**THREE CASES**

**Uncorrelated**

- $C = 0$
- Better inference about $X$, knew $Y$
- Better infer $X, Y$

**Positively Correlated**

- $Y + mx = \text{const}$
- Better infer $Y - mx$
- $Y - x$

**Negatively Correlated**

- $Y + mx = \text{const}$
- Better infer $Y + mx$
- $Y + x$
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

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

Approximation

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

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

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

also

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

Solve for \( f'(x) \)

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

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

Expanding out further we can find

\[ f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + \mathcal{O}(h^2) \]
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,

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

\[
\frac{\partial^2 L}{\partial x \partial y} \bigg|_{x_0, y_0} = \frac{L(x_0 + h, y_0 + h) - L(x_0 - h, y_0 + h) - L(x_0 + h, y_0 - h) + L(x_0 - h, y_0 - h)}{4h^2} + O(h^3)
\]

How big should \( h \) be?

The computer truncation error is about \( O(\varepsilon) \)

where \( \varepsilon \) is the smallest machine number

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

For one:

\[
\frac{O(\varepsilon)}{h^2} \sim O(h^2) \quad \text{Formula truncation error}
\]

For two of them:

\[
\left( \frac{O(\varepsilon)}{h} \right)^2 = h^3 \Rightarrow h \sim O(\varepsilon^{2/3})
\]

For \( O(\varepsilon) \approx 1 \times 10^{-6} \Rightarrow h \approx 3 \times 10^{-4} \)

For \( O(\varepsilon) \approx 2 \times 10^{-208} \Rightarrow h \approx 4 \times 10^{-69} \)

In general for:

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

\[
\frac{\partial^2 L}{\partial x_i \partial x_j} \quad \text{vary } x_i \text{ and } x_j \text{ only as in eqn above for } x + y.
\]