Two-Dimensional Problems

Consider a model with two parameters: x, y

The posterior probability is:

$$P(x, y \mid D, I)$$

The solution x_o, y_o will be given be the solution to the two simultaneous equations

$$\left.\frac{\partial P}{\partial x}\right|_{x_o, y_o} = 0 \qquad \left.\frac{\partial P}{\partial y}\right|_{x_o, y_o} = 0$$

or equivalently

$$\left.\frac{\partial \log P}{\partial x}\right|_{x_o, y_o} = 0 \qquad \left.\frac{\partial \log P}{\partial y}\right|_{x_o, y_o} = 0$$

This will give two equations

(a) $f_x(x_o, y_o) = 0$

(b) $f_y(x_o, y_o) = 0$

Strategies

1. Solve for x_o and y_o analytically ⇒ ANALYTIC SOLN
2. Solve for x_o and y_o by iterating ⇒ ITERATIVE FIXED POINT SOLN
   Guess x_o, solve (a) for y_o, solve (b) for x_o and repeat
   DOESN'T ALWAYS WORK
3. Solve for x_o and y_o numerically

Uncertainty in Two-D Problems

Again, we take a Taylor series expansion

$$L = \operatorname{Log} P(x, y \mid D, I)$$

$$L = L(x_0, y_0) + \cancelto{0}{\left.\frac{\partial L}{\partial x}\right|_{x_0,y_0}}(x - x_0) + \cancelto{0}{\left.\frac{\partial L}{\partial y}\right|_{x_0,y_0}}(y - y_0) +$$

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

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

Recall that $\dfrac{\partial^2 L}{\partial x\,\partial y} = \dfrac{\partial^2 L}{\partial y\,\partial x}$

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

$$A = \left.\frac{\partial^2 L}{\partial x^2}\right|_{x_0,y_0} \qquad B = \left.\frac{\partial^2 L}{\partial y^2}\right|_{x_0,y_0} \qquad C = \left.\frac{\partial^2 L}{\partial x\,\partial y}\right|_{x_0,y_0}$$

Taking the Exponential

$$P(x, y \mid D, I) = \exp L \propto \exp\left[-\frac{1}{2}\begin{pmatrix} x - x_0 & y - y_0 \end{pmatrix}^T \underbrace{\begin{pmatrix} A & C \\ C & B \end{pmatrix}}_{\vec{\nabla}\vec{\nabla} L}\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}\right]$$

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

Hessian
Matrix of 2nd Partial Derivs

Uncertainty in Two-D Problems

To find the uncertainty in our estimate of X.

we first marginalize out Y. This will give us

a posterior probability X.

$$P(X \mid D, I) = \int_{-\infty}^{+\infty} P(X, Y \mid D, I)\, dY$$

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

Let $x' = x - x_0 \quad dx' = dx$

$y' = y - y_0 \quad dy' = dy$

$$= K \int_{-\infty}^{\infty} \mathrm{Exp}\left[\tfrac{1}{2}\left(Ax'^2 + By'^2 + 2Cx'y'\right)\right] dy'$$

$$= K\, \mathrm{Exp}\left[\tfrac{1}{2}Ax'^2\right] \int_{-\infty}^{\infty} \mathrm{Exp}\left[\tfrac{1}{2}\left(By'^2 + 2Cx'y'\right)\right] dy'$$

Complete the square

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

$$= K\,\mathrm{Exp}\left[\tfrac{1}{2}Ax'^2\right] \mathrm{Exp}\left[-\frac{1}{2}\frac{C^2}{B}x^2\right] \int_{-\infty}^{\infty} \mathrm{Exp}\left[\tfrac{1}{2}B\left(Y' + \frac{Cx'}{B}\right)^2\right] dy'$$

Uncertainty in Two-D Problems

$$P(x\mid D,I) = K\ \mathrm{Exp}\left[\left(\frac{A}{2} - \frac{1}{2}\frac{C^2}{B}\right)x'^2\right] \int_{-\infty}^{\infty} \mathrm{Exp}\left[\frac{1}{2}B\left(y' + \frac{Cx'}{B}\right)^2\right] dy'$$

let $u = y' + \dfrac{Cx'}{B}$ $\quad du = dy'$

$$= K\ \mathrm{Exp}\left[\left(\frac{A}{2} - \frac{1}{2}\frac{C^2}{B}\right)x'^2\right] \int_{-\infty}^{\infty} \mathrm{Exp}\left[\frac{1}{2}B u^2\right] du$$

let $\sigma^2 = -\dfrac{1}{B}$ then

$$= K\ \mathrm{Exp}\left[\left(\frac{A}{2} - \frac{1}{2}\frac{C^2}{B}\right)x'^2\right] \underbrace{\int_{-\infty}^{\infty} \mathrm{Exp}\left[-\frac{u^2}{2\sigma^2}\right] du}_{\sigma\sqrt{2\pi}}$$

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

$$\propto \mathrm{Exp}\left[-\frac{1}{2}\underbrace{\left(\frac{AB - C^2}{-B}\right)}_{\equiv\ \frac{1}{\sigma_x^2}}(X - x_0)^2\right]$$

$$\Rightarrow \sigma_x = \sqrt{\frac{-B}{AB - C^2}} \qquad \text{Similarly} \qquad \sigma_y = \sqrt{\frac{-A}{AB - C^2}}$$

Uncertainty in Two-D Problems

Consider the Variance of X

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

we already did the integral over Y

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

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

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

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

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 $\pm \tan^{-1}\sqrt{\frac{A}{B}}$ wrt 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

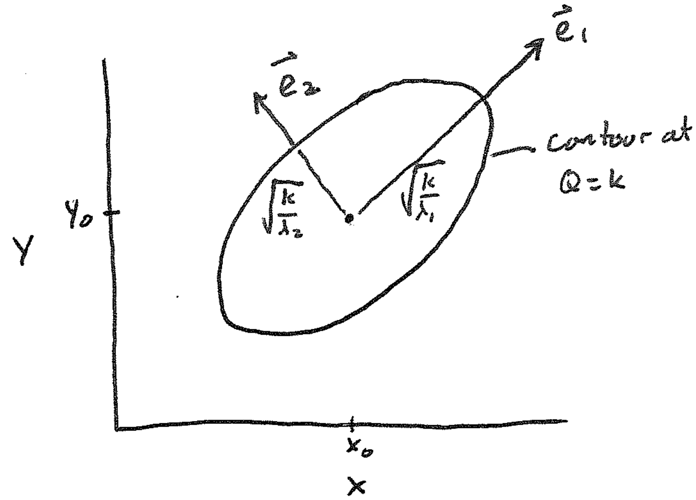
$$\text{COV} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = -\begin{pmatrix} A & C \\ C & B \end{pmatrix}^{-1} = -(\nabla\nabla L)^{-1}$$

$$= -\begin{pmatrix} \left.\dfrac{\partial^2 \operatorname{Log} P}{\partial x^2}\right|_{x_0,y_0} & \left.\dfrac{\partial^2 \operatorname{Log} P}{\partial x\,\partial y}\right|_{x_0,y_0} \\ \left.\dfrac{\partial^2 \operatorname{Log} P}{\partial y\,\partial x}\right|_{x_0,y_0} & \left.\dfrac{\partial^2 \operatorname{Log} P}{\partial y^2}\right|_{x_0,y_0} \end{pmatrix}^{-1}$$

$$= -\underbrace{H}_{\text{Hessian}}^{-1}$$

Looking more closely at our quadratic approx...

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



where λ_1 and λ_2 are the eigenvalues of H and $\vec{e}_1$ and $\vec{e}_2$ are the eigen vectors.

$$\begin{pmatrix} A & C \\ C & B \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda\begin{pmatrix} x \\ y \end{pmatrix}$$

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

$\Rightarrow A < 0, \; B < 0$ 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 $C = \dfrac{\partial^2 \operatorname{Log} P}{\partial x \, \partial y}$

For this reason, we can't just take

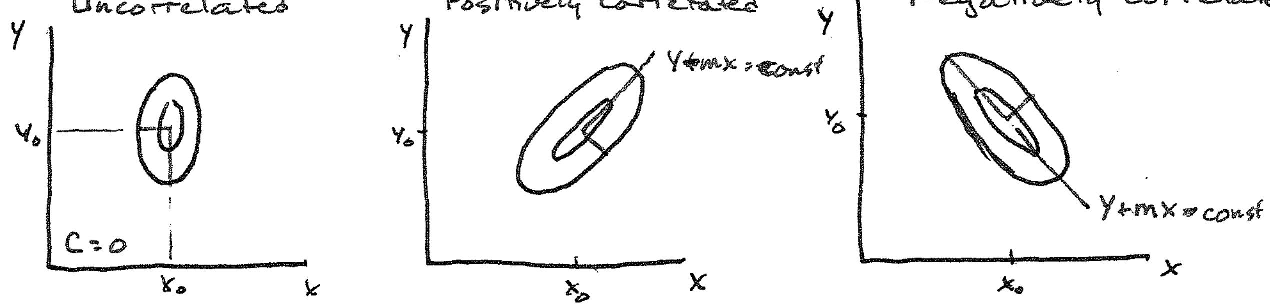
$$\sigma_x^2 \neq -\left(\frac{\partial^2 \operatorname{Log} P}{\partial x^2}\right)^{-1}$$

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

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

THREE CASES

Uncorrelated | Positively Correlated | Negatively Correlated



Better inferences about x, than y

Infer x, y

Better infer $y - mx$

$y - x$

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

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

Approximation

$$f'(x_i) = \frac{f(x_i+h) - f(x_i)}{h}$$

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

$$f(x_i+h) \simeq f(x_i) + h f'(x_i) + \frac{h^2}{2} f''(x_i) + O(h^3)$$

also

$$f(x_i-h) \simeq f(x_i) + h f'(x_i) + \frac{h^2}{2} f''(x_i) + O(h^3)$$

Solve for $f'(x_i)$

$$f(x_i+h) - f(x_i-h) \simeq + 2h f'(x_i)$$

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

Expanding out further we can find

$$f''(x_i) = \frac{f(x_i+h) - 2f(x_i) + f(x_i-h)}{h^2} + O(h^2)$$

Approximating the Hessian : 2D

$$\left.\frac{\partial^2 L}{\partial x^2}\right|_{x_0, y_0} = \frac{L(x_0+h, y_0) - 2L(x_0, y_0) + L(x_0-h, y_0)}{h^2}$$

Similarly

$$\left.\frac{\partial^2 L}{\partial y^2}\right|_{x_0, y_0} = \frac{L(x_0, y_0+h) - 2L(x_0, y_0) + L(x_0, y_0-h)}{h^2}$$

$$\left.\frac{\partial^2 L}{\partial x \partial y}\right|_{x_0, y_0} = \frac{L(x_0+h, y_0+h) - L(x_0-h, y_0+h) - L(x_0+h, y_0-h) + L(x_0-h, y_0-1}{4h^2} + \mathcal{O}(h^{2}$$

How big should h be?

The computer truncation error is about $\mathcal{O}(\epsilon)$ where ϵ is the smallest machine number (as long as L is not too complicated)

FOR ONE FINITE DIFFERENCE $\frac{\mathcal{O}(\epsilon)}{h^2} \sim \mathcal{O}(h^3)$ ~ Formula truncation error

FOR TWO OF THEM $\frac{\left(\frac{\mathcal{O}(\epsilon)}{h}\right)^2}{h} = h^3 \Rightarrow h \sim \mathcal{O}(\epsilon^{2/9})$

FOR $\mathcal{O}(\epsilon) \sim 1 \times 10^{-16} \Rightarrow h \sim 3\times 10^{-4}$

FOR $\mathcal{O}(\epsilon) \sim 2\times 10^{-308}$ MATLAB $\Rightarrow h \sim 4\times 10^{-69}$

In general for:

$\left.\frac{\partial^2 L}{\partial x_i^2}\right|_{\bar{x}_0}$ vary x_i only keep all other x_j $(j \neq i)$ constant at x_{0j}

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