6/13/19

Th1a-1

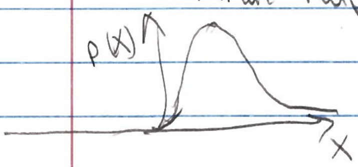
Loose Ends

(0) Discussion questions
(1) Go over information for mini-project.

(2) Recap of signal + background notebook.

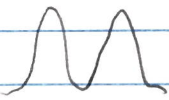
(3) Error propagation for multivariate Gaussians
(4) MLE in linear algebra form

Comments on Wednesday sampling session:
- We're very happy with interactions and interesting questions that you came up with ⇒ keep doing it!
- We'll try to share via discussion questions ⇒ put good answers
- New to us as well: e.g. asymmetric proposal distribution. What happens? One group used Maxwell → $p(x) \propto x^2 e^{-x^2}$



Only steps in one direction, so piles up at the boundary ⇒ not ergodic; meaning it does not sample the full accessible phase space.

- Picking a second random number in Poisson, why not reuse the one used to decide which way to go?

- What goes wrong with distributions like  ? Seem to get overshoot.

- Why do autocorrelation functions have exponential decay?

Th1a-2

6/13/19

Let's consider the general problem of linear combinations of multivariate Gaussian-distributed variables.

Do the easy problems first: add two Gaussian random variables, each drawn from a normal distribution but with different means and widths and independent of each other:

$$X|\mu_x,\sigma_x \sim N(\mu_x,\sigma_x^2) \Leftrightarrow p(x|\mu_x,\sigma_x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2}$$

and $Y|\mu_y,\sigma_y \sim N(\mu_y,\sigma_y^2)$

Goal: figure out how $Z = X+Y$ is distributed. $\Rightarrow p(Z|I)$

- We can think of this as an <u>error propagation</u> problem (or maybe better described as <u>uncertainty propagation</u>). How do errors combine?

- Class: What do you do? Recall CLT proof: same plan!

① $p(z) = \int_{-\infty}^{\infty} dx\, dy\, \underset{\text{joint probability}}{p(z,x,y)}$ marginalization

② $= \int_{-\infty}^{\infty} dx\, dy\, p(z|x,y)\, p(x,y)$ product rule

③ $= \int_{-\infty}^{\infty} dx\, dy\, p(z|x,y)\, p(x)\, p(y)$ independence

④ $p(z|x,y) = \delta(z-x-y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dw\, e^{-iw(z-(x+y))} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dw\, e^{-iwz} e^{iwx} e^{iwy}$

⑤ $\Rightarrow p(z) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(\int_{-\infty}^{\infty} \frac{dw}{2\pi} e^{-iwz}\right) e^{iwx} p(x)\, e^{iwy} p(y)$

$$= \int_{-\infty}^{\infty} \frac{dw}{2\pi} e^{-iwz} \left[\int_{-\infty}^{\infty} dx\, e^{iwx} p(x)\right]\left[\int_{-\infty}^{\infty} dy\, e^{iwy} p(y)\right]$$

FT of a Gaussian: $\int_{-\infty}^{\infty} dx\, e^{iwx} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} = e^{i\mu_x w} e^{-\frac{\sigma_x^2 w^2}{2}}$

$$\int_{-\infty}^{\infty} dy\, e^{iwy} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} = e^{i\mu_y w} e^{-\frac{\sigma_y^2 w^2}{2}}$$

$$\int_{-\infty}^{\infty} e^{\pm i a w} e^{-bw^2} dw = \sqrt{\frac{\pi}{b}} e^{-\frac{a^2}{4b}}$$

Th IG-3

6/13/19

$$\Rightarrow p(z) = \int_{-\infty}^{\infty} \frac{dw}{2\pi} e^{-iw(z-(\mu_x+\mu_y))} e^{-\frac{w^2}{2}(\sigma_x^2+\sigma_y^2)}$$

$$= \frac{1}{2\pi}\sqrt{\frac{2\pi}{\sigma_x^2+\sigma_y^2}} e^{-\frac{(z-(\mu_x+\mu_y))^2}{\frac{4}{2}\cdot(\sigma_x^2+\sigma_y^2)}}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_x^2+\sigma_y^2}} e^{-\frac{(z-(\mu_x+\mu_y))^2}{2(\sigma_x^2+\sigma_y^2)}}$$

$$\Rightarrow X+Y \sim N(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$$

So means add, and variances add.

Generalizations:

1 $X-Y \Rightarrow$ step 1 has $\delta(z-(x-y))$ and the minus sign carries through all the way, to 5, where $e^{iwy} \to e^{-iwy}$ and the FT becomes $e^{-i\mu_y w} \Rightarrow \mu_x-\mu_y$, otherwise unchanged.

2 $aX+bY$ for any a,b.
$\Rightarrow$ step 4 $\delta(z-(ax+by)) \Rightarrow e^{-iwz} e^{iawx} e^{ibwy}$

3 $X_1+X_2+\dots+X_m \Rightarrow$ all the same steps $\Rightarrow \sim N(\mu_{x_1}+\mu_{x_2}+\dots\mu_{x_m}, \sigma_1^2+\sigma_2^2+\dots+\sigma_m^2)$

$$\int_{-\infty}^{\infty} dx\, e^{iawx} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} = e^{ia\mu_x w} e^{-\frac{1}{2}a^2\sigma_x^2 w^2}$$

so means we can take $\mu_x \to a\mu_x$ and $\sigma_x^2 \to a^2\sigma_x^2$ everywhere

$$\Rightarrow aX+bY \sim N(a\mu_x+b\mu_y, a^2\sigma_x^2+b^2\sigma_y^2)$$

Note that 1 is a special case of 2.

6/13/19

But what if X and Y are jointly normally distributed random variables? This means $p(x,y) \neq p(x)\,p(y)$

But $p(x,y) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}(x-\mu_x \;\; y-\mu_y)\underbrace{\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}^{-1}}_{\Sigma}\begin{pmatrix}x-\mu_x \\ y-\mu_y\end{pmatrix}}$

So we have to redo the proof from $p(z) = \int_{-\infty}^{\infty} dx\,dy\; \underbrace{p(z|x,y)}_{\delta(z-(x+y))}\, p(x,y)$

Easiest perhaps to just brute force the integral (eg. Mathematica)

Result is that $aX + bY \sim N(a\mu_x + b\mu_y,\; a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y)$

- Note that if $\rho<0$ with $ab>0$ or if $\rho>0$ with $ab<0$, the variance is <u>reduced</u>.

- Special case of equal variance for convenience. What about $X-Y$?

$$\Rightarrow X-Y \sim N(\mu_x-\mu_y,\; 2\sigma^2(1-\rho))$$

- If $\rho=0$, reduces to adding in quadrature.
- but if $\rho \simeq 1$ then error bar is reduced by factor $\sqrt{1-\rho}$!
(If $\sigma_x \gg \sigma_y$, one is still dominated by the variance σ_x^2 and the correlation does not help)

6/13/19

- Where do we see such correlations?
  One place is in the energy spectrum from calculation of nuclei. Look at recent paper by Pieter Maris from June 9

- Calculations of excitation energies seem to be much more robust that those of individual energy levels. in NCSM (Hamiltonian matrix diagonalization)
  $\Rightarrow$ correlated!

- But how can we estimate ρ?
  One possibility: use theoretical calculations in different harmonic oscillator bases $\Rightarrow$ vary N_{max} and $\hbar\Omega$.
  - Compare the energies as scatter plot $\Rightarrow$ strong correlation (see figure)
  - Use the plot to estimate ρ $\Rightarrow$ argue that dependence on basis is same physics.

- Note: all unpublished so far!

6/13/19

How do we generalize to the multivariate case, where

$$X \sim N(\vec{\mu}_x, \Sigma_x) \text{ and } Y \sim N(\vec{\mu}_y, \Sigma_y) ?$$

covariance matrices

Claim: $AX + BY \sim N(A\vec{\mu}_x + B\vec{\mu}_y, A\Sigma_x A^T + B\Sigma_y B^T)$

where A, B are known $m \times N$

X, Y are independent of length N $\quad m \times N \Rightarrow m \quad (m\times N)\cdot(N\times N)\cdot(N\times m) \Rightarrow m\cdot m$.

Simpler for $A \to a$, $B \to b$:

$$\Rightarrow aX + bY \sim N(a\vec{\mu}_x + b\vec{\mu}_y, a^2\Sigma_x + b^2\Sigma_y)$$

$$\Rightarrow X + Y \sim N(\vec{\mu}_x + \vec{\mu}_y, \Sigma_x + \Sigma_y)$$
$$X - Y \sim N(\vec{\mu}_x - \vec{\mu}_y, \Sigma_x + \Sigma_y)$$

Here if $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

and so on for higher order, the normal distribution is

$$\rho(x_1, x_2, \dots x_n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu})^T \Sigma^{-1} (\vec{X}-\vec{\mu})}$$

(a) Check that it reduces to familiar case if $\rho = 0$:

Then $|\Sigma| = \det\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} = \sigma_1^2\sigma_2^2 \Rightarrow \rho(x_1, x_2) \to \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(x_1-\mu_1)\frac{1}{\sigma_1^2}(x_1-\mu_1)}$

$\times \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}(x_2-\mu_2)\frac{1}{\sigma_2^2}(x_2-\mu_2)}$ ✓

6/13/19

Data analysis recipes: Fitting a model to data

Hogg, Bovy, Lang, arXiv: 1008.4686
- provocative statements about fitting straight lines
  "Let us break with tradition and observe that in almost all cases in which scientists fit a straight line to their data, they are doing something simultaneously wrong and unnecessary."
  - wrong because very rare that a set of two dimensional measurements are truly drawn from a narrow, linear relationship
  - probably not linear in detail
  - unnecessary because communicated slope and intercept is much less informative than the full distribution of data.
- 41 extended reference notes.
- Take a look!

- In the body of the paper many interesting things, we'll extract one basic result: slope and intercept when we have correlated data

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \Rightarrow N\times1 \text{ matrix} \quad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} N\times2 \text{ matrix} \quad \Sigma = \begin{bmatrix} \sigma_{y_1}^2 & \rho_{12}\sigma_{y_1}\sigma_{y_2} & \cdots \\ \rho_{12}\sigma_{y_1}\sigma_{y_2} & \sigma_{y_2}^2 & \\ \vdots & & \ddots \\ & & & \sigma_{y_N}^2 \end{bmatrix} N\times N \text{ matrix}$$

We want to find $\Theta = \begin{bmatrix} b \\ m \end{bmatrix}$ 2x1 matrix where $\underset{N\times1}{Y} = A\,\Theta$ $\quad (N\times2)\cdot(2\times1) = 2\times1$ ✓

Check N=2

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} b+mx_1 \\ b+mx_2 \end{bmatrix} \text{ so each row works ✓}$$

- For the reader: convince yourself N=3 or higher is correct

6/13/19

Now why don't we just solve $Y = A\theta \Rightarrow \theta = A^{-1}Y$?

- This equation is over constrained.

- Actual frequentist answer comes from minimizing χ^2 (maximum likelihood)

$$\chi^2 = \sum_{i=1}^{N} \frac{[y_i - f(x_i)]^2}{\sigma_{y_i}^2} \overset{\text{more general}}{=} [Y - A\theta]^T \Sigma^{-1} [Y - A\theta]$$

if uncorrelated

Claim: result is (can you prove? see next page)

$$\theta = \begin{bmatrix} b \\ m \end{bmatrix} = [\underbrace{A^T \Sigma^{-1} A}_{2\times 2}]^{-1} [A^T \Sigma^{-1} Y]$$

$2\times N \quad N\times N \quad N\times 1$

Why can't I say $[A^T\Sigma^{-1}A]^{-1} = A^{-1}\Sigma A^{T-1} A^T \Sigma^{-1} Y = A^{-1}Y$?

Plausibility of expression for θ. We need square, invertible matrices

$$Y = A\theta \qquad N\times 1 = (N\times 2)\cdot(2\times 1) \checkmark$$

$$\Sigma^{-1}Y = \Sigma^{-1}A\theta \qquad (N\times N)\cdot(N\times 1) = (N\times N)(N\times 2)(2\times 1) \checkmark$$

$$A^T\Sigma^{-1}Y = A^T\Sigma^{-1}A\theta \qquad (2\times N)(N\times N)(N\times 1) = \underbrace{(2\times N)(N\times N)(N\times 2)}_{2\times 2}(2\times 1)$$

now we can invert $\Rightarrow \theta = (A^T\Sigma^{-1}A)^{-1}(A^T\Sigma^{-1}Y)$

Problem: what if $y_i = q x_i^2 + m x_i + b$? $\Rightarrow \theta = \begin{bmatrix} b \\ m \\ q \end{bmatrix}$

What is A now? $\begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix}$

- How does this work with Bayesian? Same as before, include priors, now just matrix valued.
  - We'll see this next week.

Th1a-9

6/13/19

$$\chi^2 = (Y - A\theta)^T \Sigma^{-1} (Y - A\theta) = (Y_i - A_{ij}\theta_j)\Sigma^{-1}_{ii'}(Y_{i'} - A_{i'j'}\theta_{j'})$$

$$\Rightarrow \frac{\partial \chi^2}{\partial k} = -A_{ij}\delta_{jk}\Sigma^{-1}_{ii'}(Y_{i'} - A_{i'j'}\theta_{j'}) + (Y_i - A_{ij}\theta_j)\Sigma^{-1}_{ii'}(-A_{i'j'}\delta_{j'k})$$
$$= 0$$

Move θ to the opposite side and show doubled terms are equal (overall multiply by -1)

$$A_{ik}\Sigma^{-1}_{ii'}Y_{i'} + Y_i\Sigma^{-1}_{ii'}A_{i'k} = A_{ik}\Sigma^{-1}_{ii'}A_{i'j'}\theta_{j'} + A_{ij}\theta_j\Sigma^{-1}_{ii'}A_{i'k}$$

$\overbrace{\quad}^{j' \to j}$ ← symmetric

$$\underbrace{A_{ik}}_{A^T_{ki}}\Sigma^{-1}_{ii'}Y_{i'} + A_{i'k}\Sigma^{-1}_{ii'}Y_i = A_{ik}\Sigma^{-1}_{ii'}A_{i'j}\theta_j + A_{i'k}\Sigma^{-1}_{i'i}A_{ij}\theta_j$$

$i \leftrightarrow i'$ and $\Sigma^{-1}_{i'i} = \Sigma^{-1}_{ii'}$ 　　 same with $i \leftrightarrow i'$

$$\Rightarrow 2(A^T)_{ki}\Sigma^{-1}_{ii'}Y_{i'} = 2(A^T)_{ki}(\Sigma^{-1})_{ii'}A_{i'j}\theta_j$$

← square, invertible

$$\text{or} \quad (A^T\Sigma Y) = (A^T\Sigma^{-1}A)\theta$$

$$\Rightarrow \theta = (A^T\Sigma^{-1}A)^{-1}(A^T\Sigma Y) \qquad \text{QED}$$