9/25/19

Physics 8805: Learning from Data: Lecture 10

Notebooks for today:
- visualization_of_CLT.ipynb
- parameter_estimation_fitting_straight_line_II.ipynb
- MCMC-diagnostics.ipynb

- Follow up to Assignment 2 proof of $\lim_{D\to\infty} \frac{D^N e^{-D}}{N!} = \frac{1}{\sqrt{2\pi D}} e^{-(N-D)^2/2D}$ $\quad N \equiv D(1+\delta)$

  $D\delta^2/2$ is kept but $\delta/2$ dropped. Why? Ans: $N-D \sim \sqrt{D}$ ← standard deviation
  $\Rightarrow D\delta \sim \sqrt{D} \gg 1$.

- Start with visualization_of_CLT.ipynb notebook.

- Follow-up to where we left off on Friday. Quick recap:
  - looking a parameter_estimation_fitting_straight_line_II.ipynb
  - extracting posterior for Hubble constant from data for velocities of galaxies vs. distance.
  - don't care about intercept ⇒ nuisance parameter (therefore marginalize)
  - comparison to maximum likelihood estimate
    ⇒ in this case not too different unless the best intercept is used, leading to a too-small estimate.

- Using emcee to do the sampling
  ⇒ (56) and marginalization with MCMC means just neglecting columns for marginalized parameters.

  another argument ⇒ • Recall $\frac{\int g(\theta_0,\theta_1) f(\theta_0,\theta_1)\, d\theta_0 d\theta_1}{\int f(\theta_0,\theta_1)\, d\theta_0 d\theta_1} \doteq \sum_{i=1}^{N} g(\theta_0^{(i)}, \theta_1^{(i)})$

  but suppose g doesn't depend on θ_0? Then the sum doesn't care about the θ_0 column in $\vec{\theta}^{(i)}$! It just uses the θ_1 column.

- Continue with Error propagation in that notebook.
  - See how including the full posterior lead to larger (more realistic) uncertainty.

9/25/19

PDF Manipulations I: Combining random variables

- Suppose we have two random variables, X and Y, drawn from two different distributions.
  - For concreteness we'll take the distributions to be Gaussian (normal) with different means and widths, and independent of each other.

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

$$Y|\mu_y,\sigma_y \sim N(\mu_y,\sigma_y^2) \Leftrightarrow p(y|\mu_y,\sigma_y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2}$$

- What is the distribution of the sum of X and Y?
  - From past experience, we might expect that "errors add in quadrature", but how does that arise in detail?
  - This is a special case of uncertainty propagation: How do errors combine?

- First phrase this in term of a statement about posteriors: Goal: given $Z=X+Y$, how is Z distributed, ie what is $p(Z|I)$? Plan: follow the same steps as with the CLT proof!

Drop I here

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 (what if not the case?)

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

Now we can substitute ④ into ③ and the integrals factorize.
Note: still general for independent pdfs for X,Y (haven't used Gaussian).

9/25/19 (62)

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

need to verify that we can change integration orders

$$= \int_{-\infty}^{\infty} \frac{dw}{2\pi} e^{-iwz} \underbrace{\left[\int_{-\infty}^{\infty} dx\, e^{iwx} p(x) \right] \left[\int_{-\infty}^{\infty} dy\, e^{iwy} p(y) \right]}_{\text{Fourier transforms (FTs) of pdf's}}$$

Finally we get specific

⑥ FT of a Gaussian: $\int_{-\infty}^{\infty} dx\, e^{iwx} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} = e^{i\mu_x w} e^{-\frac{\sigma_x^2 w^2}{2}}$

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

Here we make $\int_{-\infty}^{\infty} e^{\pm iaw} e^{-bw^2} dw = \sqrt{\frac{\pi}{b}} e^{-a^2/4b}$

$$⑦ \overset{⑤+⑥}{\Rightarrow} p(z) = \int_{-\infty}^{\infty} \frac{dw}{2\pi} e^{-iw(z-(\mu_x+\mu_y))} e^{-\frac{w^2}{2}(\sigma_x^2+\sigma_y^2)}$$

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

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

$\Rightarrow X+Y \sim N(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$: means add, <u>variances add</u>

Generalizations:

Ⓐ $X-Y \Rightarrow$ step ④ becomes $\delta(z-(x-y))$ and the minus sign carries through all the way to ⑥, where $e^{iwy} \to e^{-iwy}$ and the FT becomes $e^{-i\mu_y w} \Rightarrow N(\mu_x-\mu_y, \sigma_x^2+\sigma_y^2)$

Ⓑ $aX+bY$ as any $a,b \Rightarrow$ ④ becomes $\delta(z-(ax+by)) \Rightarrow e^{-iwz} e^{iawx} e^{ibwy}$

$$\Rightarrow \int_{-\infty}^{\infty} dx\, e^{iawx} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} = \underset{\mu_x \to a\mu_x}{e^{i(a\mu_x)w}}\; \underset{\sigma_x^2 \to a^2\sigma_x^2}{e^{-\frac{1}{2}a^2\sigma_x^2 w^2}} \Rightarrow aX+bY \sim N(a\mu_x+b\mu_y, a\sigma_x^2+b\sigma_y^2)$$

Ⓒ $X_1+X_2+\ldots+X_m \Rightarrow$ all the same steps!

$$\Rightarrow \sim N(\mu_1+\mu_2+\ldots+\mu_m, \sigma_1^2+\sigma_2^2+\ldots+\sigma_m^2)$$

9/25/19

* Note that given the FTs of the pdfs being combined, we can generalize ⑥ and ⑦ from the case of normal distributions.

· Now suppose that X and Y are jointly normally distributed random variables, but not independent. So $p(x,y) \neq p(x)\,p(y)$.

· General form:

(note inverse)

$$p(x,y) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}\left(x-\mu_x \;\; y-\mu_y\right)\Sigma^{-1}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}}$$

← determinant of Σ

$$\boxed{\equiv \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})} \quad \vec{x} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \vec{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}}$$

where we parametrize the symmetric, positive definite matrix (i.e., positive eigenvalues) by

"covariance matrix" ⇒ $$\Sigma \equiv \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \qquad 0 \leq \rho \leq 1$$

If $\rho = 0$, then we reduce to the uncorrelated case already considered.

Returning to the derivation, we redo it from $p(z) = \int_{-\infty}^{\infty} dx\,dy\; p(z|x,y)\,p(x,y)$, with $p(z|x,y) = \delta(z-(x+y))$

This is left as an exercise for the reader (or Mathematica!):

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

- So if $\rho \approx 1$, $\Rightarrow N(a\mu_x + b\mu_y,\; (a\sigma_x + b\sigma_y)^2)$
- Note that if $\rho < 0$ with $a \cdot b > 0$ or $\rho > 0$ with $ab < 0$, then the variance is <u>reduced</u> compared to adding in quadrature.

· Special case of equal variance for convenience: $X - Y \sim N(\mu_x - \mu_y,\; 2\sigma^2(1-\rho))$
- If $\rho = 0$, back to adding in quadrature
- But if $\rho \approx 1$, then error bar reduced by factor $\sqrt{1-\rho}$!

(If $\sigma_x \gg \sigma_y$, still dominated by σ_x^2 and correlation doesn't affect.)

9/25/19 (64)

Ok, now one further generalization: X and Y are themselves from (independent) multivariate normal distributions of dimension N while A, B are known M×N matrices. So

$$X \sim N(\underset{\text{N vector}}{\vec{\mu}_x}, \underset{\text{N×N covariance matrices}}{\Sigma_x}) \quad \text{and} \quad Y \sim N(\vec{\mu}_y, \Sigma_y)$$

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

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

and so on for high order, then the normal distribution is

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

- check that it works for $N=2$, $\rho=0$ ✓

- If we apply this for $A \to a\underset{\text{N×N identity}}{I_N}$, $B \to bI_N$ (so M=N here, with a, b scalars) then it is simpler:

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

so the <u>covariance</u> matrices add in quadrature.