6/11/19 Bayes 2019 TALENT Lecture T1b

Plan: Extend exploration of pdfs, go back to straight line, do notebook on "amplitude vs. background" problem

- Return to Exploring_pdfs.ipynb
  - Two-dimensional pdfs → review QM analogies at top of notebook
  - Recall that β distribution as $a = b \to$ large is Gaussian with same mean and standard deviation

    [come back to two variable posteriors below]
    - line 34 → mu2 = beta2.dist.mean()
      sigma2 = beta2.dist.std()
    - try a2=b2=2,5,10,20,50
  - Then try t distribution with same insertions
    - Add norm1_dist = stats.normal(0.,1.)
      ax1.plot(x,t, norm1_dist.pdf(x,t), cor='red')
      and same for ax2, ax3.

    ⇒ as $\nu \to \infty$, t distribution approaches standard normal distribution. ← $\mu = 0$, $\sigma = 1$
  - These are realizations of the central limit theorem (CLT)
  - *** M1b-8 ⇒ Before proving CLT, do M1b-8 ← also done in T1a.
- Most general form of CLT: the sum of n random values drawn from any pdf of finite variance σ^2 tends as $n \to \infty$ to be Gaussian distributed about the expectation value of the sum, with variance $n\sigma^2$.

- Consequences:
  1. The mean of a large numbers of values becomes normally distributed <u>regardless</u> of the probability distribution from which the values are drawn.

  2. Functions such as the Binomial and Poisson distribution all tend to look like Gaussian distributions in the limit of a large number of drawings.

     E.g. $P_n = \frac{\lambda^n e^{-\lambda}}{n!}$ (n integer) $\xrightarrow[n\to\infty]{} p(x) = \frac{e^{-(x-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}}$

6/11/19

independent random variables

Suppose $x_1, \ldots, x_n$ drawn from a distribution with mean $\langle x \rangle = \int x\, p(x)\, dx = 0$ and $\langle x^2 \rangle = \sigma^2$

Let $X = \frac{1}{\sqrt{n}}(x_1 + x_2 + \ldots x_n) = \sum_{j=1}^{n} \frac{x_j}{\sqrt{n}}$ need to scale by $\frac{1}{\sqrt{n}}$ for finite X

What is the distribution of X [call it $p(X|I)$]?

Class: fill in the blank to justify the following standard steps:

Suppress I here

$$p(X) = \int_{-\infty}^{\infty} dx_1 \ldots dx_n \; p(X, x_1, \ldots, x_n) \quad \text{marginalization}$$ why?

$$= \int_{-\infty}^{\infty} dx_1 \ldots dx_n \; p(X|x_1, \ldots, x_n)\, p(x_1, \ldots x_n) \quad \text{product rule}$$

$$= \int_{-\infty}^{\infty} dx_1 \ldots dx_n \; p(X|x_1, \ldots, x_n)\, p(x_1)\, p(x_2) \ldots p(x_n) \quad \text{independence}$$

What is $p(X|x_1, \ldots, x_n)$? $\Rightarrow = \delta(X - \frac{1}{\sqrt{n}}(x_1 + \ldots x_n))$

Rather than use it to evaluate one of the integrals, use a fourier representation:

$$\delta(X - \tfrac{1}{\sqrt{n}}(x_1 + \ldots + x_n)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \; e^{-i\omega(X - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i)} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \; e^{-i\omega X} \prod_{i=1}^{n} \left[\int_{-\infty}^{\infty} dx_i \, e^{\frac{i\omega x_i}{\sqrt{n}}} p(x_i)\right]$$

but the [ ] terms are all the same! Suppose we Taylor expand $e^{\frac{i\omega x_i}{\sqrt{n}}}$
(Fourier integral dominated by small x as $n \to \infty$; when does this fail.)

$$e^{\frac{i\omega x}{\sqrt{n}}} = 1 + \frac{i\omega x}{\sqrt{n}} + \frac{(i\omega)^2 x^2}{2n} + O\left(\frac{\omega^3 x^3}{n^{3/2}}\right) \Rightarrow \int_{-\infty}^{\infty} dx\, p(x)\left[1 + \frac{i\omega x}{\sqrt{n}} + \frac{(i\omega)^2 x^2}{2n} + \ldots\right]$$

$$\Rightarrow = 1 + \frac{i\omega}{\sqrt{n}}\underbrace{\cancel{\langle x \rangle}}_{0} - \frac{\omega^2}{2n}\underbrace{\cancel{\langle x^2 \rangle}}_{\sigma^2} + \langle x^3 \rangle O\left(\frac{\omega^3}{n^{3/2}}\right)$$ ← σ^2 assumed to be finite

$$\Rightarrow p(X) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \; e^{-i\omega X}\left[1 - \frac{\omega^2}{n}\sigma^2 + O\left(\frac{\omega^3}{n^{3/2}}\right)\right]^n$$ but $\lim_{n\to\infty}(1 + \frac{a}{n})^n = e^a$

$$\Rightarrow = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \; e^{-i\omega X} e^{-\omega^2 \sigma^2} = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{X^2}{2\sigma^2}} \quad \text{QED}$$

- generalize to $\langle x \rangle \neq 0$ with $X = x_1 + \ldots x_n - n\mu$ and change to $y_i = x_i - \mu$.
- Why does it work? product of many pdfs smooths it out and kills fat tails.

6/11/19

- Rest of Exploring_pdfs.ipynb has 2d projected posteriors
  - Check behavior. Are the distributions correlated?
    No: signature is no "tilt" to distribution → see T1b-4
  - Review what is being plotted. ⇒ do M1b-9 for definition of confidence intervals
  - Comment on "multi-modal" distribution. "modes" are peaks, so multi-modal has more than one.

Now back to parameter_estimation_fitting_straight_line_I.ipynb
- do M1b-7

Comments on notebook:
- note that x_i is also randomly distributed uniformly
- log likelihood gives fluctuating results "whose size depend on # of data points N and standard deviation of noise dy.
  ⇒ if time, explore in exercise session how "size" varies with these.

intercept $b = 25.0$
slope $m = 0.5$
$\sigma = 5.0$

- Compare priors on slope ⇒ uniform in m vs. uniform in angle
- implementation of plots comparing priors – class comments
  - with first set of data with N=20 points, does prior matter? No!
  - with second set " " " N=3 points " " " ? Yes!
- note log posterior = log(likelihood) + log(prior)
  - maximum taken to be 1 for plotting
  - exponentiate: posterior = exp(log_posterior)

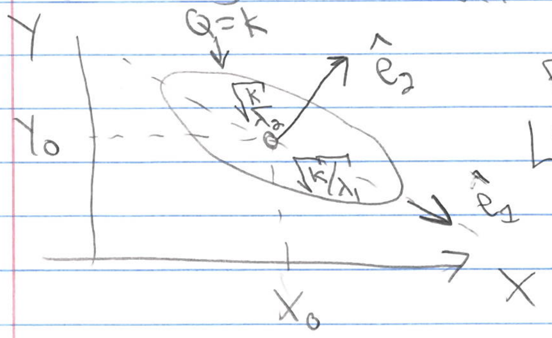
¥ • What does it mean that the ellipses are slanted? see T1b-4

- 2nd set of data: Flat gives $b = -50 \pm 75$, $m = 1.5 \pm 1$ so barely in 1σ
  symmetric gives $b = 25 \pm 50$, $m = .5 \pm .75$ so much better!

Switch to corner!

6/11/19

Likelihoods with two variables (or posteriors) with quadratic approximation



Find X_0, Y_0 (best estimate) by differentiating
$L(X,Y) = \log p(X,Y \mid \{data\}, I)$

$$\Rightarrow \left.\frac{\partial L}{\partial X}\right|_{X_0,Y_0} = 0 \qquad \left.\frac{\partial L}{\partial Y}\right|_{X_0,Y_0} = 0$$

- To check reliability, Taylor expand around $L(X_0, Y_0)$:

$$L = L(X_0,Y_0) + \frac{1}{2}\left[\left.\frac{\partial^2 L}{\partial X^2}\right|_{X_0,Y_0}(X-X_0)^2 + \left.\frac{\partial^2 L}{\partial Y^2}\right|_{X_0,Y_0}(Y-Y_0)^2 + 2\left.\frac{\partial^2 L}{\partial X \partial Y}\right|_{X_0,Y_0}(X-X_0)(Y-Y_0) + \dots\right] + \dots \equiv L(X_0,Y_0) + \frac{1}{2}Q + \dots$$

Makes sense to do this in matrix notation

$$Q = \begin{pmatrix} X-X_0 & Y-Y_0 \end{pmatrix} \begin{pmatrix} A & C \\ C & B \end{pmatrix} \begin{pmatrix} X-X_0 \\ Y-Y_0 \end{pmatrix}$$ symmetric

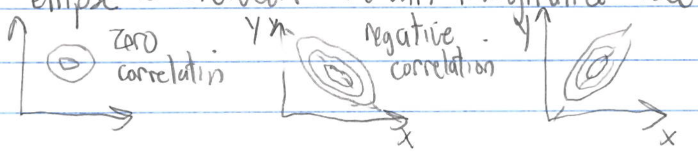
$$A = \left.\frac{\partial^2 L}{\partial X^2}\right|_{X_0,Y_0} \qquad B = \left.\frac{\partial^2 L}{\partial Y^2}\right|_{X_0,Y_0} \qquad C = \left.\frac{\partial^2 L}{\partial X \partial Y}\right|_{X_0,Y_0}$$

- So in quadratic approximation, the contour $Q=k$ for some k is an ellipse centered at X_0, Y_0. Orientation and eccentricity determined by A, B, and C.

- Principal axes found from eigenvectors of $\begin{pmatrix} A & C \\ C & B \end{pmatrix}$ (Hessian matrix)

$$\begin{pmatrix} A & C \\ C & B \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \lambda_1, \lambda_2 < 0 \text{ (so } X_0, Y_0 \text{ is a maximum)}$$

What if ellipse is skewed? We can marginalize (see Wednesday lecture)

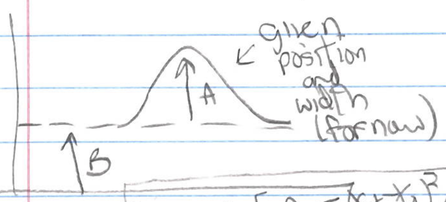


Look at correlation matrix

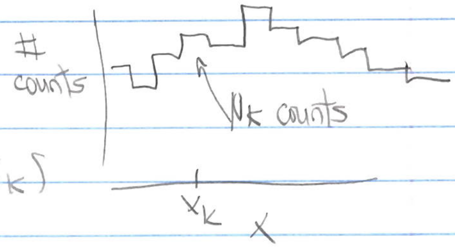
$$\begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix} = -\begin{pmatrix} A & C \\ C & B \end{pmatrix}^{-1}$$ (more on Thursday)

6/11/19

Siva example: Amplitude of a signal in the presence of background



$\Rightarrow \{N_k\}$ counts in bins $\{x_k\}$



n_0 scales with time counts collected

$$\boxed{D_k = n_0\left[A\, e^{-(x_k - x_0)^2/2w^2} + B\right]}$$

Goal: Given $\{N_k\}$ find A, B

- So what is the posterior we want?

  $p(A, B \mid \{N_k\}, I)$ with $I = x_0, w$, Gaussian, flat background

Actual counts we get will be integers, and we can expect a Poisson distribution

$$\Rightarrow p(n \mid \mu) = \frac{\mu^n e^{-\mu}}{n!} \quad \text{for } n \geq 0 \text{ integer}$$

or

$$n \to N_k,\ \mu \to D_k \qquad p(N_k \mid D_k) = \frac{D_k^{N_k} e^{-D_k}}{N_k!} \quad \text{for } k^{th} \text{ bin at } x_k$$

* What do we learn from the plots of the Poisson distribution?

$$\underset{\text{posterior}}{p(A,B \mid \{N_k\}, I)} \propto \underset{\text{likelihood}}{p(\{N_k\} \mid A, B, I)} \times \underset{\text{prior}}{p(A, B \mid I)}$$

$$\Rightarrow L = \log\left[p(A,B \mid \{N_k\}, I)\right] = \text{constant} + \sum_{k=1}^{M}\left[N_k \log(D_k) - D_k\right]$$

- Choose constant for convenience: independent of A, B.
- Best point estimate: maximize $L(A,B)$ to find A_0, B_0.

$$p(A,B \mid I) = \begin{cases} \text{constant} & 0 \leq A \leq A_{max},\ 0 \leq B \leq B_{max} \\ 0 & \text{otherwise} \end{cases}$$

- Look at code for likelihood and prior
- Uniform flat prior for $0 \leq A \leq A_{max}$, $0 < B \leq B_{max}$
  - Not sensitive to A_{max}, B_{max} if larger than support of likelihood

Fig #	data bins	ΔX	$(X_k)_{max}$	D_{max}
1	15	1	7	100
2	15	1	7	10
3	31	1	15	100
4	7	1	3	100

T1b-6

6/11/19

Comments on Figures.

Fig 1: 15 bins and $D_{max} = 100$

- Contours are at 20% intervals showing height.
- Read off best estimates and compare to True.
  - does find signal is about half background
- Marginalization of B
  - What if we don't care about B? "nuisance parameter"

$$p(A|\{N_k\}, I) = \int_0^\infty p(A,B|\{N_k\}, I)\, dB$$

compare to $p(A|\{N_k\}, B_{true}, I) \Rightarrow$ plotted on graph

- Also can marginalize over A

$$p(B|\{N_k\}, I) = \int_0^\infty p(A,B|\{N_k\}, I)\, dB$$

- See how these are done in code; B_marginalized
  B_true_fixed
  - note the normalization at the end.

- Set extra plots to true
  - different representations of same info and contours in first 3. Last one is attempt at 68%, 95%, 99.7% but looks wrong.

* - note difference between contours showing pdf height and showing integrated volume.

- Look at the other Figures and draw conclusions

- How should you design your experiments.