LECTURE I - PROBABILITY

John Gough
Aberystwyth

CIRM, 16-20 April 2018

1.1 Jelly Beans

		Colour			
		G(reen)	Y(ellow)	B(lue)	
Texture	R(ough)	**10**	**40**	**0**	50
	S(mooth)	**20**	**10**	**20**	50
		30	50	20	100

- Joint Probability - Prob{G; R} = 0.10, etc.
- Marginal Probability – Prob{G} = 0.30, Prob{R} = 0.50
- Conditional Probability – Prob{G|R} = 0.20
- Are Colour & Texture independent variables?
- How could they be made so?

1.2 Probability Density Functions (PDFs)

- PDFs

A random variable X has pdf ρ_X so that

$$\Pr\left\{x \le X < x + dx\right\} = \rho_X\left(x\right)\, dx.$$

Normalization requires $\int \rho_X\left(x\right) dx = 1$.

- Joint PDFs

A pair of random variables X, Y has joint pdf $\rho_{X,Y}$ so that

$$\Pr\left\{x \le X < x + dx; y \le Y \le y + dy\right\} = \rho_{X,Y}\left(x, y\right)\, dxdy.$$

Normalization requires $\int \rho_{X,Y}\left(x, y\right) dxdy = 1$.

1.2 Probability Density Functions (PDFs)

- Marginal PDFs

$$\begin{aligned} \rho_X(x) &= \int \rho_{X,Y}(x,y)\,dy, \quad (x\text{—marginal}) \\ \rho_Y(y) &= \int \rho_{X,Y}(x,y)\,dx, \quad (y\text{—marginal}) \end{aligned}$$

- Conditional PDFs

The pdf for X given that $Y = y$ is defined to be

$$\rho_X(x|y) = \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}.$$

1.2 Probability Density Functions (PDFs)

- Statistical Independence

We say that X and Y are **statistically independent** if their joint probability factors into the marginals

$$\rho_{X,Y}(x,y) = \rho_X(x) \times \rho_Y(y), \quad \text{(independence)}.$$

- Consequence

In the special case where X and Y are independent we have

$$\rho_X(x|y) = \rho_X(x).$$

In other words, conditioning on the fact that $Y = y$ makes no change to our knowledge of X.

1.3 Bayes Theorem

Given $\rho_{Y|X}$ and ρ_X we can work out $\rho_{X|Y}$.

The key is that we can work out the joint pdf:

$$\rho_{X,Y}(x,y) = \rho_{Y|X}(y|x)\,\rho_X(x),$$

and so

$$\rho_{X|Y}(x|y) = \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} = \frac{\rho_{Y|X}(y|x)\,\rho_X(x)}{\int \rho_{Y|X}(y|x')\,\rho_X(x')\,dx'}.$$

1.4 Estimation

- We have a variable, *X*, which is unknown (even its pdf!)
- We measure a second variable, Y, somehow dependent on *X*.
- Wish to estimate *X* from the measurement of *Y.*

- Likelihood …

Our main modelling assumption is that whenever X is known to take a particular value of x, then the conditional pdf for Y is a known function: we write this as

$$\lambda\left(y|x\right).$$

- But we want things the other way round! *X* is unknown, not *Y*!
- We need the marginal for *X* to use Bayes Theorem, but we don't know it.
- Therefore we guess an **a prior distribution** for *X*:

$$\rho_X(x) = \rho_{\text{prior}}(x).$$

- We then have the corresponding joint probability for *X* and *Y* :

$$\rho^{\text{prior}}_{X,Y}(x,y) = \lambda(y|x) \times \rho_{\text{prior}}(x).$$

- If we subsequently measure *Y* = *y* then we obtain the **a posteriori distribution**:

$$\rho_{\text{post}}(x|y) = \frac{\lambda(y|x)\,\rho_{\text{prior}}(x)}{\int \lambda(y|x')\,\rho_{\text{prior}}(x')\,dx'}$$

Example (Signal + Noise)

- Let X be the position of a particle. We measure

$$Y = X + \sigma Z$$

where Z is a standard normal variable, called the "noise", independent of X.

- The likelihood function is

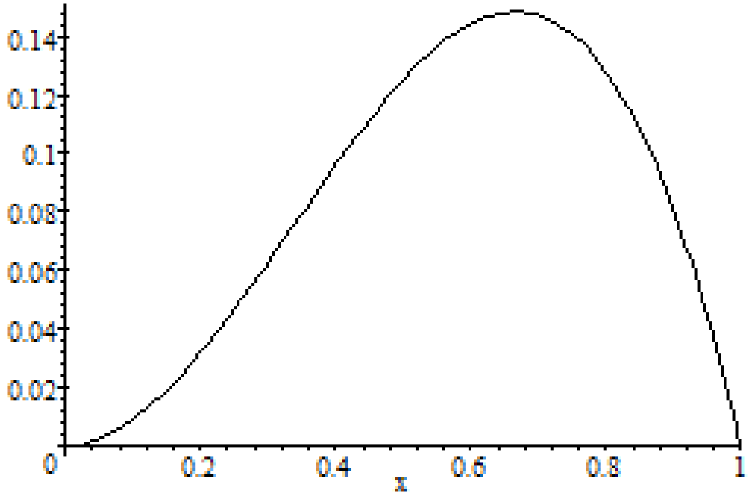
$$\lambda\left(y|x\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-x)^2/2\sigma^2},$$

- If we choose a prior ρ_{prior} for X then

$$\rho_{\text{post}}\left(x|y\right) = \frac{\rho_{\text{prior}}\left(x\right) e^{-(y-x)^2/2\sigma^2}}{\int \rho_{\text{prior}}\left(x'\right) e^{-(y-x')^2/2\sigma^2} dx'}.$$

Example (Parameter Estimation)

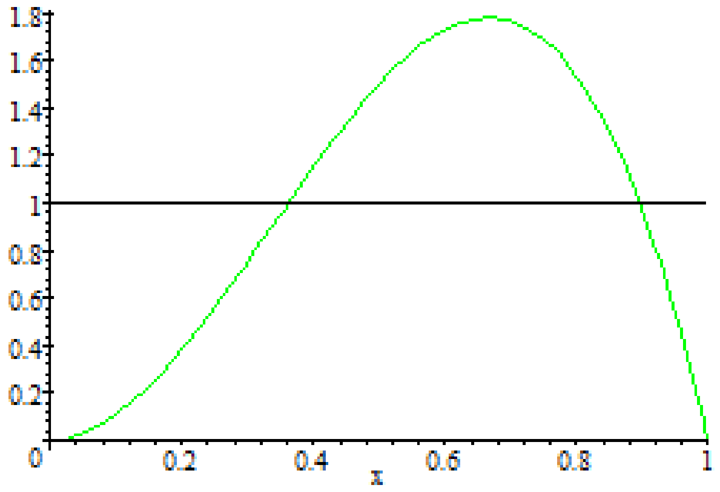
- Suppose we have a coin with an unknown probability, *x*, for heads.
- We toss it three times and obtain the sequence *y* = HHT.
- The likelihood function is then $\lambda(HHT|x) = x^2(1-x)\,, \quad 0 \le x \le 1.$



Choose a Prior

Let us choose the prior to be the uniform distribution $\rho_X(x) = 1$, that is, we take all values for the probability parameter x to be equally likely. A simple calculation gives

$$\rho_{\text{post}}(x|HHT) = \frac{x^2(x-1)}{\int_0^1 x'^2(1-x')\,dx'} = 12x^2(1-x)\,.$$



The *a posteriori* distribution has

Mean 3/5

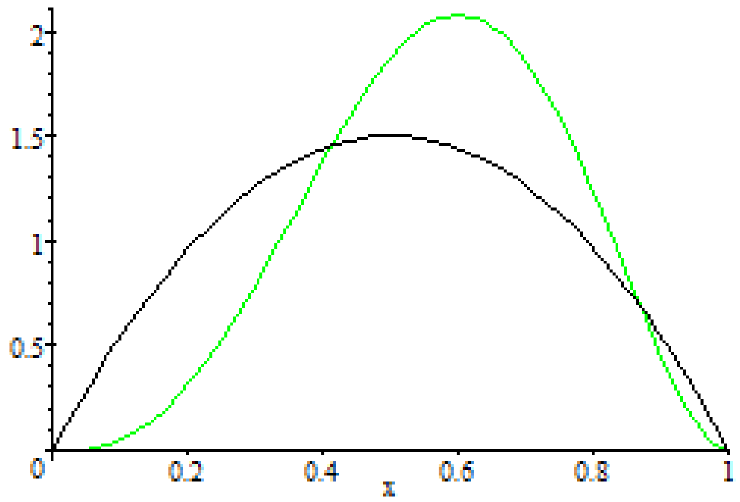
Mode 2/3

If we had however chosen a different prior, we would get a different answer. For instance, if we set

$$\rho_{\text{prior}}(x) = 6x(1-x), \quad 0 \le x \le 1,$$

then we calculate

$$\rho_{\text{post}}(x|HHT) = \frac{x^3(x-1)^2}{\int_0^1 x'^3(1-x')^2\,dx'} = 60x^3(1-x)^2.$$



The *a posteriori* distribution has

Mean 4/7

Mode 3/5

Example (Von Neumann's Model for Quantum Measurement)

- We consider an observable, $\hat{X}$, say the position of a quantum system.

- Rather than measure $\hat{X}$ directly, we measure an observable $\hat{Y}$ giving the pointer position of a second system (called the measurement apparatus). We assume that apparatus is described by a wave-function ϕ.

- The initial state of the system and apparatus is $|\Psi_0\rangle = |\psi_{\text{prior}}\rangle \otimes |\phi\rangle$, i.e.,

$$\langle x, y|\Psi_0\rangle = \psi_{\text{prior}}\,(x)\;\phi\,(y)\,.$$

We couple the particle to the apparatus using the unitary

$$\hat{U} = e^{i\mu\,\hat{X}\otimes\hat{P}_Y/\hbar},$$

where $\hat{P}_Y = -i\hbar\frac{\partial}{\partial y}$.

After coupling, the joint state is

$$\langle x, y|\hat{U}\Psi_0\rangle = \psi_{\text{prior}}\,(x)\;\phi\,(y-\mu x)\,.$$

If the measured value of $\hat{Y}$ is y, then the **a posteriori wave-function** is

$$\psi_{\text{post}}(x|y) = \frac{1}{\sqrt{\rho_Y(y)}} \psi_{\text{prior}}(x)\ \phi(y - \mu x)$$

where

$$\rho_Y(y) = \int |\psi_{\text{prior}}(x)\ \phi(y - \mu x)|^2 dx.$$

Basically, the pointer position will be a random variable with pdf given by ρ_Y: the a posteriori wave-function may then be thought of as a random wave-function on the system Hilbert space:

$$\psi_{\text{prior}}(x) \longrightarrow \psi_{\text{post}}(x|Y).$$

It is interesting to reconsider the problem in the **Heisenberg picture**!

Let $\hat{Y}^{\text{in}} = I \otimes \hat{Y}$. In the Heisenberg picture, the observable that we actually measure is

$$\hat{Y}^{\text{out}} = \hat{U}^*\big(I \otimes \hat{Y}^{\text{in}}\big)\hat{U} = I \otimes \hat{Y}^{\text{in}} + \mu\,\hat{U}^*\big(\hat{X} \otimes I\big)\hat{U}.$$

- Here $\hat{Y}^{\text{out}}$ is the measured observable.
- We estimate the *signal*, $\hat{U}^*(\hat{X} \otimes I)\hat{U}$.
- And we add *noise*, $\hat{Y}^{\text{in}}$, which has pdf $|\phi(y)|^2$.