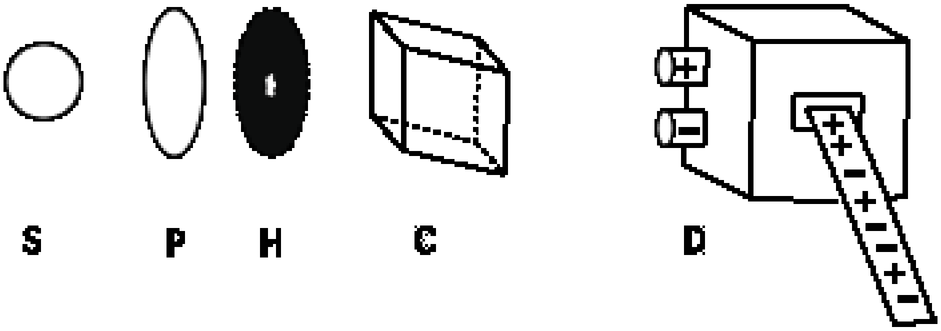
Some Bayesian Thoughts

The essence of quantum theory is its ability to predict **probabilities** for the outcomes of tests based on specified preparations. Quantum mechanics is **not** a theory about reality; it is a prescription for making the best possible predictions about the future based on certain information (specified) about the past. The quantum theorist can tell you **what the odds are**, if you wish to bet on the occurrence of various events, such as the clicking of this or that detector.

However, a more common activity is the reverse situation where the outcomes of tests are known and it is their preparation(the initial state) that has to be guessed. The formal name for this process is **retrodiction**. Retrodicting is the analog of forecasting an event, but directed oppositely in time, i.e., to the past rather than the future. Just as one might forecast, from a knowledge of physical laws along with specific data about the current position and speed of a comet, where it will be ten years from now, one might retrodict where it was ten years ago.

Suppose we have the experiment described in the figure below:



where S is a thermal source of light, P is a polarizer, H is pinhole, C is a calcite crystal, and D is a detector with separate counters for the two different polarized beams emerging from the calcite crystal. The detector D also makes a **permanent record** of the measured events. We assume that the light intensity is so weak and the detectors are so fast that individual photons can be registered. The arrivals of photons are recorded by printing + or − on a paper tape, according to whether the upper or lower detector was triggered, respectively. The sequence of + and − marks appears **random**. As the total number marks, N_+ and N_-, become large, we find that the corresponding probabilities(count ratios), tend to limits

$$\frac{N_+}{N_+ + N_-} \to \cos^2\alpha \quad , \quad \frac{N_-}{N_+ + N_-} \to \sin^2\alpha$$

where α is the angle between the polarization axis of the polaroid and the optic axis of the calcite crystal.

Now suppose that we do an experiment and find that the two detectors recorded 4 and 3 events, respectively. What can we **infer** about the orientation of the polarizer?

This is the so-called **"inverse probability"** problem, which is an ideal situation to use Bayesian methods.

Consider the following description.

Event B is the outcome of the experiment described above with 4 + detections and 3 – detections. This is a **single** experiment and **not** a set of seven experiments.

Event A is the positioning of the polarizer at an angle in the interval θ to $\theta + d\theta$, in that experiment.

Now, in a **statistical ensemble**, that is, an infinite set of conceptual replicas of the same system, the **relative frequencies** of events A and B define the probabilities $prob(A|I) = prob(A)$ and $prob(B|I) = prob(B)$, where I is all the information about the preparation(conditioning).

In addition, $prob(A\,\&\,B|I)$ is the joint probability of events A and B. This is the relative frequency of the occurrence of **both** events, in the statistical ensemble under consideration. $prob(A|B\,\&\,I)$ is the **conditional probability** of A, when B is true.

As before we have the relations

$$prob(A\,\&\,B\,|\,I) = prob(A\,|\,B\,\&\,I)prob(B\,|\,I) = prob(B\,|\,A\,\&\,I)prob(A\,|\,I)$$

and

$$prob(A\,|\,B\,\&\,I) = \frac{prob(B\,|\,A\,\&\,I)prob(A\,|\,I)}{prob(B\,|\,I)}$$

The last equation is **Baye's theorem.**

In this equation, it is assumed that $prob(B|A\,\&\,I)$ is known from the appropriate physical theory. For example, in the above experiment, the theory tells us that the probabilities for triggering the upper and lower detectors are $\cos^2\theta$ and $\sin^2\theta$. We therefore, have from the Binomial distribution (the Binomial coefficient)

$$prob(B = \{4,3\}\,|\,A\,\&\,I) = \frac{(n_+ + n_-)!}{n_+!n_-!}(prob(+\,|\,A\,\&\,I))^{n_+}(prob(-\,|\,A\,\&\,I))^{n_-}$$

$$= \frac{7!}{4!3!}(\cos^2\theta)^4(\sin^2\theta)^3 = \frac{7!}{4!3!}\cos^8\theta\sin^6\theta = 35\cos^8\theta\sin^6\theta$$

In order to determine $prob(A|B\,\&\,I)$ we still need $prob(A|I)$ and $prob(B|I)$. These probabilities cannot be calculated from a theory nor determined empirically. They depend solely on the statistical ensemble that we have mentally constructed.

Let us consider the **complete** set of events of type A and call them $A_1, A_{2,}......$ etc. For example, A_j represents the positioning of the polarizer at an angle between θ_j and $\theta_j + d\theta_j$. By completeness,

$$\sum_j P(A_j) = 1$$

and therefore,

$$P(B) = \sum_j P(B \mid A_j) P(A_j)$$

At this point we introduce Baye's postulate (different from Baye's theorem). This postulate we have used in earlier discussions is also called the **"principle of indifference"** or the **"principle of insufficient reasoning"**.

If we have no reason to expect that the person who positioned the polarizer had a preference for some particular orientation, we assume that all orientations are equally likely, so that $P(A) = \frac{d\theta}{\pi}$ for every θ (we can always take $0 \le \theta \le \pi$ because θ and $\theta + \pi$ are equivalent).

We then have

$$P(B) = \sum_j P(B \mid A_j) P(A_j) = \frac{1}{\pi} \int_0^{\pi} 35 \cos^8\theta \sin^6\theta d\theta = \frac{135}{2^{11}}$$

and we then obtain from Baye's theorem that

$$prob(A \mid B \,\&\, I) = \frac{prob(B \mid A \,\&\, I) prob(A \mid I)}{prob(B \mid I)}$$

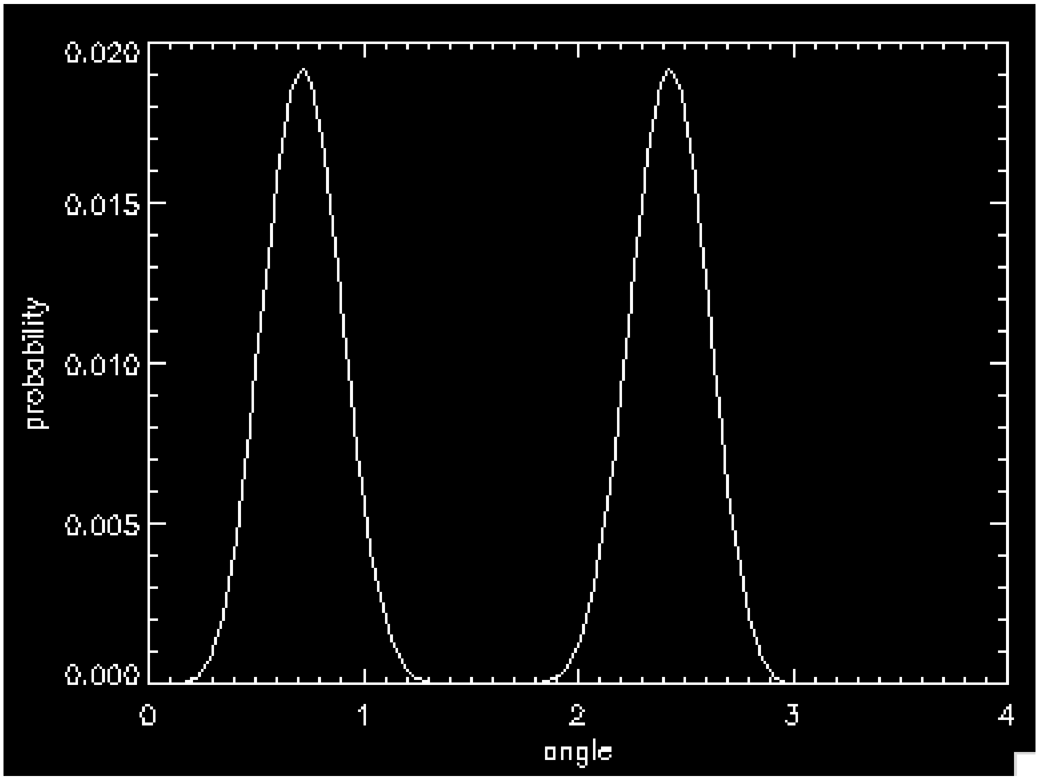
$$= \frac{\frac{7!}{4!3!} \cos^8\theta \sin^6\theta \frac{d\theta}{\pi}}{\frac{135}{2^{11}}} = \frac{2^{11}}{5\pi} \cos^8\theta \sin^6\theta d\theta$$

which is the probability that the angle is between θ and $\theta + d\theta$ given that event B is the outcome of a single experiment with 4 + detections and 3 – detections.

Suppose $d\theta = 1^\circ = 0.0175 rad$. If we plot

$$prob(\theta \mid B = \{4,3\}, d\theta = 0.0175) = \frac{2^{11}}{5\pi} \cos^8\theta \sin^6\theta d\theta = 2.283 \cos^8\theta \sin^6\theta$$

versus θ we have



This says that, given the single data set, the angle is most likely to be

$$0.72\,rad = 41.3^\circ \quad or \quad 2.42 rad = 138.7^\circ$$

Clearly, Bayesian analysis allows us to infer results from one-time experiments on single systems. The key is the use of Baye's postulate of indifference.