5 Assigning probabilities

Sum and product rule specify the relationships between pdfs. But they don't tell us how to assign them. Are there any general principles that could help us?

Bernoulli + Keynes: "principle of indifference":

If we can enumerate a set of mutually exclusive possibilities and we have no reason to believe that one is more likely than other, then we should assign same probability.

Example: Ordinary die, six outcomes $X_i \equiv$ face on top i dots

$\Rightarrow$ $\text{prob}(X_i | I) = 1/6$

But what does our statement mean? Suppose you enumerate A–F instead of X_i. Then ascribe probabilities to A–F. Now suppose that someone say that our mapping A–F $\rightarrow X_i$ was wrong, e.g. $A \Leftrightarrow X_2$ instead of $A \Leftrightarrow X_3$. Should that make any difference?

If we have no information about this, the answer is no! So <u>consistency</u> demands that probability should not change if the order is changed. Only way for this is $\text{prob}(X_i | I) = 1/6$.

$\Rightarrow$ "desideratum of consitency"

Consider urn with R red balls and W-white balls.
Principle of indifference says that

$$\mathrm{prob}(j \mid I) = \frac{1}{R+W}$$

is the probability that any ball "j" will be chosen.
Marginalize + product rule:

$$\mathrm{prob}(\mathrm{red} \mid I) = \sum_{j=1}^{R+W} \mathrm{prob}(\mathrm{red}, j \mid I)$$

$$= \sum \mathrm{prob}(j \mid I)\, \mathrm{prob}(\mathrm{red} \mid j, I)$$

$$= \frac{1}{R+W} \sum \underbrace{\mathrm{prob}(\mathrm{red} \mid j, I)}_{= \begin{cases} 0 ; & \text{ball } j \text{ is white} \\ 1 ; & \text{ball } j \text{ is red} \end{cases}}$$

$$= \frac{R}{R+W}$$

Location + scale parameters

Can we extend the principle of indifference to continuous variables?

Remember that the probability that the position of a lighthouse X lies in the range $x+\delta x$ is

$$\mathrm{prob}(X=x \mid I)\, dX = \lim_{\delta x \to 0} \mathrm{prob}(x \le X \le x+\delta x \mid I)$$

Now suppose that we made a mistake in defining the origin and that the position previously quoted as x is now $x+x_0$. Should this change the pdf assigned to X. If I indicates gross ignorance then the answer is again no.

$$\Rightarrow \quad \mathrm{prob}(X \mid I)\, dX = \mathrm{prob}(X + x_0 \mid I)\, d(X+x_0)$$

$$\Rightarrow \mathrm{prob}(X \mid I)\, dX = \mathrm{prob}(X + x_0 \mid I)\, dX$$

$$\Rightarrow \mathrm{prob}(X \mid I) = \mathrm{prob}(X + x_0 \mid I)$$

$$\Rightarrow \underline{\mathrm{prob}(X \mid I) = \mathrm{const}}\ !$$

! So complete ignorance about location parameter is represented by uniform pdf !

Another common problem are quantities associated with size or magnitude.

For example the length L of a molecule. What should prob(L|I) be, if I meant complete ignorance? If we were told that a mistake had been made about the units of the length scale quoted, e.g. ångstroms instead of nanometers, this should not alter the pdf:

$$\mathrm{prob}(L|I)\,dL = \mathrm{prob}(\beta L|I)\,d(\beta L)$$

$$\Rightarrow \mathrm{prob}(L|I)\,dL = \mathrm{prob}(\beta L|I)\,\beta\,dL$$

$$\Rightarrow \mathrm{prob}(L|I) \propto 1/L \quad \text{"Jeffreys prior"}$$

$$\Rightarrow \mathrm{prob}(L|I)\,dL = \frac{dL}{L} = d\ln L$$

$\Rightarrow$ Jeffreys prior is equivalent to a uniform pdf for $\ln L$:

$$\mathrm{prob}(\ln L|I)\,d\ln L = \text{const}\; d\ln L \qquad \text{q.e.d.}$$

5.2. Testable information, the principle of maximum entropy

We have seen that assigning the pdf hinges on the consistency argument and requires a consideration of the transformation group which characterize the given ignorance: permutation for the dice, origin shift for location and stretch for a magnitude.

Suppose a die was rolled very often. If you were only told that the average result was 4.5, what probability should you assign to the possibilities X_i = i dots? The information yields the constraint

$$\sum_{i=1}^{6} i \, \mathrm{prob}(X_i | I) = 4.5$$

Which pdf is the best to assign? Jaynes argued that we should choose the one maximizing the entropy $S = \sum_i p_i \ln p_i$

where $p_i = \mathrm{prob}(X_i | I)$, subject to $\sum p_i = 1$

$$\sum i p_i = 4.5$$

Fig 5.2

Let's get a feel why we should use S by looking at the kangaroo problem:

Information: 1/3 of all kangaroos have blue eyes and
1/3 of all kangaroos are left-handed

Question: On the basis of I, what proportion are left handed and are blue eyed?

Four possibilities:

		Left handed	
		True	false
Blue eyed	True	p_1	p_2
	False	p_3	p_4

however normalization: $p_1 + p_2 + p_3 + p_4 = 1$

and in addition: $p_1 + p_2 = p_1 + p_3 = 1/3$

$\Rightarrow$ $p_2 = 1/3 - p_1$

$p_3 = 1/3 - p_1$; $p_1 \in [0; 1/3]$

and $p_1 + \frac{1}{3} - p_1 + \frac{1}{3} - p_1 + p_4 = 1$

$\Rightarrow$ $p_4 = p_1 - \frac{2}{3} + 1 = p_1 + 1/3$

p_1	$1/3 - p_1$
$1/3 - p_1$	$1/3 + p_1$

So all pdf's with $p_1 \in [0; 1/3]$ satisfy all information given. Which is best?

Common sense say $p_1 = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$ indicating no correlation between blue eyes and left handedness seems reasonable. So is there some principle that also gives this answer?

Table ~~Fig~~ 5.1 Skilling has shown that the only function yielding this answer is the entropy S and monotonically related functions.

We can give another argument in favor of S:

The monkey argument

Suppose M distinct possibilities $\{X_i\}$ to be considered. Imagine the propositions X_i represented by equal-sized boxes and monkeys throwing pennies into them. After $N \gg M$ pennies have been thrown, we count the number of coins per X_i and get our pdf from this. If the pdf is in agreement with I, we keep it, if not we disregard it. The pdf that in the end occurred most often would be a sensible choice for prob$(\{X_i\} | I)$.

After the monkeys have distributed all N pennies, we find n_1 in box 1, n_2 in box 2 etc. The total number is constrained to

$$N = \sum_{i=1}^{M} n_i$$

$\{n_i\} \rightarrow$ pdf $\{p_i\}$ for the possibilities $\{X_i\}$

$$p_i = \frac{n_i}{N}$$

There are M^N possibilities to scatter the coins among the boxes. But many of the sequences will yield the same $\{n_i\}$. The expected frequency F with which $\{p_i\}$ will occur is therefore given by

$$F(\{p_i\}) = \frac{\text{number of ways obtaining } \{n_i\}}{M^N}$$

To compute the number of ways to obtain $\{n_i\}$, consider the following: In how many ways can n_1 coins be chosen from N? The answer is the binomial coefficient $\binom{N}{n_1} = \frac{N!}{n_1!\,(N-n_1)!}$

Then, n_2 coins can be distributed from $(N-n_1)$ remaining coins etc:

$$\binom{N}{n_1} \times \binom{N-n_1}{n_2} \times \binom{N-n_1-n_2}{n_3} \times \cdots \binom{n_M}{n_M}$$

$$\frac{N!}{n_1!\,\cancel{(N-n_1)!}} \times \frac{\cancel{(N-n_1)!}}{n_2!\,\cancel{(N-n_1-n_2)!}} \times \frac{\cancel{(N-n_1-n_2!)}}{n_3!\,\cancel{(N-n_1-n_2-n_3!)}} \times \cdots$$

$$= \frac{N!}{n_1!\,n_2!\ldots n_M!}$$

$$\Rightarrow \ln F(\{p_i\}) = -N\ln M + \ln N! - \sum_{i=1}^{M} \ln n_i!$$

use Stirlings formula: $\ln n! = n\ln n - n$

$$\Rightarrow \ln F = -N\ln M + N\ln N - \underset{\sim}{N} - \sum n_i \ln n_i + \underset{\sim}{\sum n_i}$$

$$= -N\ln M + N\ln N - \sum n_i \ln n_i$$

$$= -N\ln M + N\ln N - \sum Np_i \ln Np_i$$

$$= -N\ln M + N\ln N - \sum Np_i \ln p_i - \sum Np_i \ln N$$

$$= -N\ln M + \underset{\sim}{N\ln N} - N\sum p_i \ln p_i - \underset{\sim}{N\ln N \underbrace{\sum p_i}_{=1}}$$

$$\Rightarrow \ln F = -N \ln M - N \sum p_i \ln p_i$$

Since this is monotonically related to the frequency with which the monkeys will come up with a candidate pdf $\{p_i\}$, the assignment probability prob $(\{x_i\} | I)$ which best represents our lack of knowledge is the one with greatest $\ln F$ being consistent with our information I.
As M and N are constants, this is equivalent to the constrained maximization of

$$S = - \sum_{i=1}^{M} p_i \ln p_i$$

The measure

Suppose that we had a situation with three possibilities:

$X_{1,2}$ = face on top has 1,2 dots
X_3 = face on top has 3...6 dots.

We would then want to make the boxes the monkeys throw their coins in different sized.
So if the chance that a monkey puts a penny in box i is m_i and $\sum m_i = 1$, then some short argument leads to

$$S = - \sum_{i=1}^{M} p_i \ln \frac{p_i}{m_i}$$ "Shannon-Jaynes entropy"