

Re: How to identify flat (even) distributions?

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2008-12/msg00059.html>

- *From:* Steve555 <foursheds@xxxxxxxxxxxxxxxx>
 - *Date:* Thu, 11 Dec 2008 04:59:55 -0800 (PST)
-

On 11 Dec, 11:25, illywhacker <illywac...@xxxxxxxx> wrote:

On Dec 11, 11:57 am, Steve555 <foursh...@xxxxxxxxxxxxxxxx> wrote:

maybe the chi-squared one will be the fastest to compute.

They all involve summing over all ten score frequencies: it is hard to see that one can avoid this!

Either way, my hunch is that it would be useful, when devising a music recommendation system, to eliminate – or give a low weighting to – the scores of these people.

It is usually better to decide exactly what you are trying to achieve before trying to achieve it. Many questions of this nature are simply linked to a failure to define the goal precisely. Once the goal is defined, whether these people are 'useful' or not should be a question of calculation based on hypothesized models, not hunches. However difficult you may think it is to model human behaviour, you have no choice if you wish to pursue this type of application. You may as well do it explicitly, thereby making your assumptions explicit, for otherwise you will in any case be doing it implicitly, and your assumptions will be hidden and hence un-analysable. Your explicit models may seem ludicrously simplistic, but this is the nature of the application. Your implicit models will also be ludicrously simplistic, but you would not be forced to face up to it and admit it.

illywhacker;

Sorry, I don't understand that at all. As best as I can understand it, I think I'm doing what you suggest!

With messy problems (like recommendation systems, or as you say, modeling human behaviour) that haven't really been solved to the point of giving very reliable predictions, creative hunches are the way to

Re: How to identify flat (even) distributions?

go!

"whether these people are 'useful' or not should be a question of calculation based on hypothesized models, not hunches."

There is no predefined calculation to solve this. problem. I stated that my hunch(idea) was that some people – based on their scoring distribution – might improve (or add noise to) my prediction algorithm. I then asked the learned people here how to best judge these scoring distributions.

The next step is the calculation based on hypothesized models.

Three things are needed:

- 1) an idea
- 2) being taught the techniques needed to pursue your idea
- 3) the execution:a calculation based on hypothesized model.

I believe I'm doing these steps in the perfect order :-)

"Your explicit models may seem ludicrously simplistic, but this is the nature of the application. Your implicit models will also be ludicrously simplistic, but you would not be forced to face up to it and admit it."

Your previous 'entropy' post helped me a lot and encouraged exploration, I would never have thought of it, and maybe it will best reflect what I'm trying to measure. But, I'm sorry, that last sentence is out there!

I think it means: "It is easier to notice how simplistic a model is when it is explicit" But what's that got to do with the price of fish?

.