Profile Likelihood

W. Nash[1], B. Stone[1]

[1]University of California Los Angeles

November 3, 2017

Dice Example

- Let's consider a model with 4 parameters $\mu = (\mu_1, \mu_2, \mu_3, \mu_4)$ which predicts how often a face will be rolled from a 4 sided die.
- The parameters of our model will be simply the probability of each face being rolled in one throw

$$\begin{array}{ll} P(1|\mu) = \mu_1 & P(3|\mu) = \mu_3 \\ P(2|\mu) = \mu_2 & P(4|\mu) = \mu_4 \end{array}$$

- For collecting our data, we simply roll the die N number of times, and find n_1 total times a 1 is rolled, etc. So we have

$$N = \sum_{i=1}^{4} n_i \qquad 1 = \sum_{i=1}^{4} \mu_i \tag{1}$$

The Likelihood Function

- Without any data, the probability for getting n_1 1's, n_2 2's and n_3 3's and n_4 4's in N total throws is given by the multinomial distribution

$$P(\mathbf{n}|\mu, N) = \frac{N!}{n_1!n_2!n_3!n_4!}\mu_1^{n_1}\mu_2^{n_2}\mu_3^{n_3}\mu_4^{n_4}$$

  Where $\mathbf{n} = (n_1, n_2, n_3, n_4)$.

- In general, if we have made an observation, we can consider its **Likelihood** as being a function of our model's parameters, instead a function of the data we observed

$$L(\mu) = P(\mathbf{n}|\mu, N)$$

- In our case, the different n's are what we observe (fixed) giving us a Likelihood function of

$$L(\mu) = \frac{N!}{n_1!n_2!n_3!n_4!}\mu_1^{n_1}\mu_2^{n_2}\mu_3^{n_3}\mu_4^{n_4}$$

We can look at the logarithm of the likelihood function since it gives us the same answer as otherwise and we only care about differences.

$$F = -2 \ln L = -2\Big[\ln\Big(\frac{N!}{n_1!n_2!n_3!n_4!}\Big) + \sum_{i=1}^{4} n_i \ln \mu_i\Big] \tag{2}$$

Since we have a constraint that the probabilities must add to 1, we can use Lagrange multipliers to solve for the maximally likely $\hat{\mu}_i$

$$F = -2\Big[\ln\Big(\frac{N!}{n_1!n_2!n_3!n_4!}\Big) + \sum_{i=1}^{4} n_i \ln \mu_i\Big] + \alpha\Big[\sum_{i=1}^{4} \mu_i - 1\Big] \tag{3}$$

Taking the derivative with respect to μ_i, and solving for α in the standard way[1], we find the maximum likelihood estimates $\hat{\mu}_i$, for $i = 1, 2, 3, 4$ to be

$$\hat{\mu}_i = \frac{n_i}{N} \tag{4}$$

[1] in backup

Let's suppose that we only care about μ_1, and take the rest as nuisance parameters. We let $\hat{\hat{\mu}}_i$ for $i \neq 1$ be the ML estimate given μ_1. In order to find the confidence interval, we need to find

$$\lambda = L(\mu_1, \hat{\hat{\mu}}_{i\neq 1})/L(\hat{\mu}) \tag{5}$$

Where λ is called the **likelihood ratio**. A quick note on the notation:

Vector	Description
μ	The parameter(s) of interest, we can freely change these
$\hat{\mu}$	The maximum likelihood estimate for the parameter(s), coordinate of the bottom of the "well"
$\hat{\hat{\mu}}_{i\neq 1}$	The nuisance parameters that gives us the maximum likelihood constrained to a value of μ_1

We already have $L(\hat{\mu})$ from just plugging in equation (4), but now we need to find $\hat{\hat{\mu}}_{i\neq 1}$. We do it in roughly the same way, but only maximize with respect to some of the parameters[2], we find that

$$\lambda = \left(\frac{\mu_1}{\hat{\mu}_1}\right)^{n_1}\left(\frac{1-\mu_1}{1-\hat{\mu}_1}\right)^{N-n_1} \tag{6}$$

So we have that our **profile likelihood function**

$$-2\ln\lambda = -2\left[n_1\ln\frac{\mu_1}{\hat{\mu}_1} + (N-n_1)\ln\left(\frac{1-\mu_1}{1-\hat{\mu}_1}\right)\right] \tag{7}$$

Where our maximum likelihood estimate is

$$\hat{\mu}_1 = \frac{n_1}{N} \tag{8}$$

[2] in backup

Plotting the Likelihood distribution

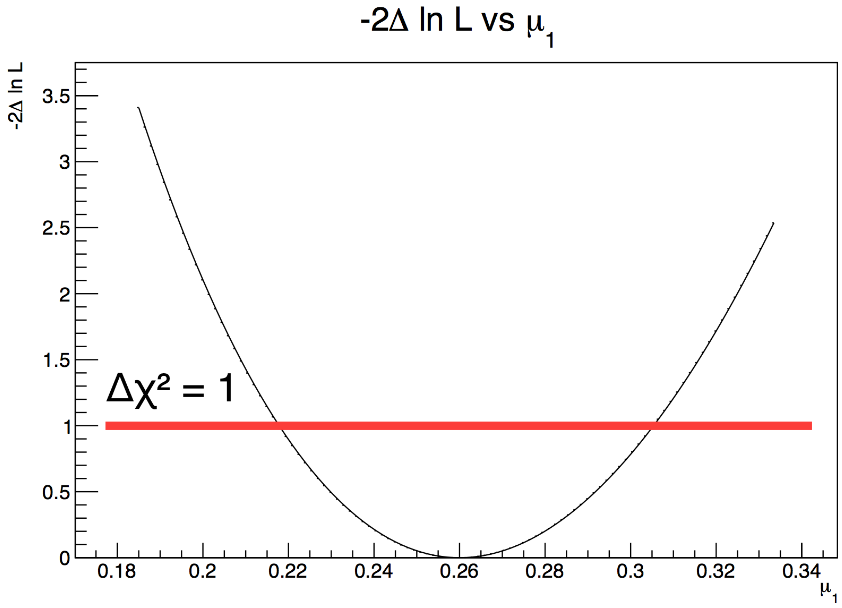
We are now fully prepped to draw the likelihood well and find the confidence interval. I wrote some code that randomly rolled $N = 100$ dice, giving

$$n_1 = 26 \tag{9}$$
$$n_2 = 27 \tag{10}$$
$$n_3 = 24 \tag{11}$$
$$n_4 = 23 \tag{12}$$

I then wrote some code to plotted the log-likelihood ratio we found with

$$-2\ln\lambda = -2\left[26\ln\frac{\mu_1}{0.26} + 74\ln\left(\frac{1-\mu_1}{0.74}\right)\right] \tag{13}$$

Moving in small steps around the maximum likelihood estimate of $\hat{\mu}_1 = 0.26$, we can find how the curve looks



-2Δ ln L vs μ1
-2Δ ln L
3.5
3
2.5
2
1.5
1
0.5
0
0.18
0.2
0.22
0.24
0.26
0.28
0.3
0.32
0.34
μ1
Δχ² = 1

```
mu1 = 0.2165      chi2 = 1.06685      mu1 = 0.302       chi2 = 0.86104
mu1 = 0.218       chi2 = 0.991428     mu1 = 0.3035      chi2 = 0.921795
mu1 = 0.2195      chi2 = 0.919015     mu1 = 0.305       chi2 = 0.984507
mu1 = 0.221       chi2 = 0.849577     mu1 = 0.3065      chi2 = 1.04917
mu1 = 0.2225      chi2 = 0.783084     mu1 = 0.308       chi2 = 1.11576
```

Looking at the numerical values in the plot, we can determine our 68% confidence interval for μ_1 to be

$$[0.218, 0.305] \tag{14}$$

This tells us that for any of the μ_1's within the interval, our data would fall within the 68% acceptance interval of μ_1 (approximately by Wilk's theorem).

Backup

Equation (4) derivation: The function we want to maximize is

$$F = -2\Big[\ln\Big(\frac{N!}{n_1!n_2!n_3!n_4!}\Big) + \sum_{i=1}^{4} n_i \ln \mu_i\Big] + \alpha\Big[\sum_{i=1}^{4}\mu_i - 1\Big] \tag{15}$$

We do this by taking the derivative with respect to one of the μ_i's

$$\frac{\partial F}{\partial \mu_i} = \frac{n_i}{\mu_i} + \alpha = 0 \rightarrow \boxed{\hat{\mu}_i = -\frac{n_i}{\alpha}} \tag{16}$$

This tells us the **maximum likelihood estimate** for our parameter μ_i, but we still don't know α, so we to plug these back into our constraint equation. In the case that all the parameters are maximized, we have

$$0 = \sum_{i=1}^{4}\hat{\mu}_i - 1 = -\frac{n_1 + n_2 + n_3 + n_4}{\alpha} - 1 \tag{17}$$

This tells us $\alpha = -N$ and therefore

$$\boxed{\hat{\mu}_i = \frac{n_i}{N}} \tag{18}$$

Equation (6) derivation: We are looking for $\hat{\hat{\mu}}_{i\neq 1}$ which by definition maximizes the log likelihood function for a given μ_1. Similar to how we found equation (4), we want to maximize

$$F = -2\Big[\ln\Big(\frac{N!}{n_1!n_2!n_3!n_4!}\Big) + \sum_{i=1}^{4} n_i \ln \mu_i\Big] + \beta\Big[\sum_{i=1}^{4}\mu_i - 1\Big] \tag{19}$$

Where we have a different Lagrange multipler because we are not maximizing for each parameter. We find

$$\frac{\partial F}{\partial \mu_{i\neq 1}} = 0 = \frac{n_{i\neq 1}}{\mu_{i\neq 1}} + \beta \rightarrow \boxed{\hat{\hat{\mu}}_{i\neq 1} = -\frac{n_{i\neq 1}}{\beta}} \tag{20}$$

Now lets plug this into our constraint to find β

$$0 = \mu_1 - 1 + \sum_{i=2}^{4}\mu_i = \mu_1 - 1 - \frac{n_2 + n_3 + n_4}{\beta} \tag{21}$$

This tells us that β is a function of μ_1, with

$$\beta = -\frac{N - n_1}{1 - \mu_1} \tag{22}$$

Now with our β, we can write that the maximally likely $\hat{\hat{\mu}}_{i\neq1}$ given any μ_1 is simply

$$\hat{\hat{\mu}}_{i\neq1}(\mu_1) = \frac{n_i(1-\mu_1)}{N-n_1} \tag{23}$$

Now we just have to plug these into our Likelihood function to find the maximally likely outcome given any μ_1

$$L(\mu_1, \hat{\hat{\mu}}_{i\neq1}) = \frac{N!}{n_1!n_2!n_3!n_4!}\mu_1^{n_1}\hat{\hat{\mu}}_2^{n_2}\hat{\hat{\mu}}_3^{n_3}\hat{\hat{\mu}}_4^{n_4} \tag{24}$$

Looking at the ratio and plugging in, the form simplifies to identically the binomial distribution!

$$\lambda = \frac{L(\mu_1, \hat{\hat{\mu}}_{i\neq1})}{L(\hat{\mu})} = \left(\frac{\mu_1}{\hat{\mu}_1}\right)^{n_1}\left(\frac{1-\mu_1}{1-\hat{\mu}_1}\right)^{N-n_1} \tag{25}$$