

# Eliciting a Dirichlet Distribution

*Jeremy Oakley*

*2019-03-26*

We illustrate the process of eliciting a Dirichlet distribution using the methodology and case study in Zapata-Vazquez, R., O'Hagan, A. and Bastos, L. S. (2014). Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics* 41, 1919-1933. Quoting from their Section 3.4,

This example concerns the efficacy of a new antibiotic in patients who are hospitalised in the Pediatric Intensive Care Unit (PICU) and who are severely infected by pneumococci (which is associated with pneumonia, meningitis, and septicaemia, among other conditions). The possible results after the infection are: to survive in good condition, to have a sequel, or to die. An expert is asked to provide judgements about the proportions of patients who will have each of these possible results. Denoting these proportions by  $\pi_1$ ,  $\pi_2$ ,  $\pi_3$ , these form a set of proportions that must sum to 1.

The Dirichlet distribution is parameterised by

$$f(\pi_1, \pi_2, \pi_3) \propto \prod_{i=1}^3 \pi_i^{d_i-1},$$

with  $n = \sum_{i=1}^3 d_i$ .

The elicited judgements for the three marginal proportions were

	Good outcome	Sequel	Dead
Lower quartile	0.50	0.22	0.11
Median	0.55	0.30	0.15
Upper quartile	0.60	0.35	0.20

For each marginal proportion  $\pi_i$ , the expert has provided a lower quartile, a median and an upper quartile, so we define a single vector of probabilities, specifying which quantiles have been elicited.

```
p1 <- c(0.25, 0.5, 0.75)
```

We then define one vector for each marginal proportion, giving the values of the elicited quantiles.

```
v.good <- c(0.5, 0.55, 0.6)
v.seq1 <- c(0.22, 0.3, 0.35)
```

```
v.dead <- c(0.11, 0.15, 0.2)
```

Next we fit probability distributions to each set of elicited quantiles.

```
library(SHELF)
fit.good <- fitdist(vals = v.good, probs = p1, lower = 0, upper = 1)
fit.seql <- fitdist(vals = v.seql, probs = p1, lower = 0, upper = 1)
fit.dead <- fitdist(vals = v.dead, probs = p1, lower = 0, upper = 1)
```

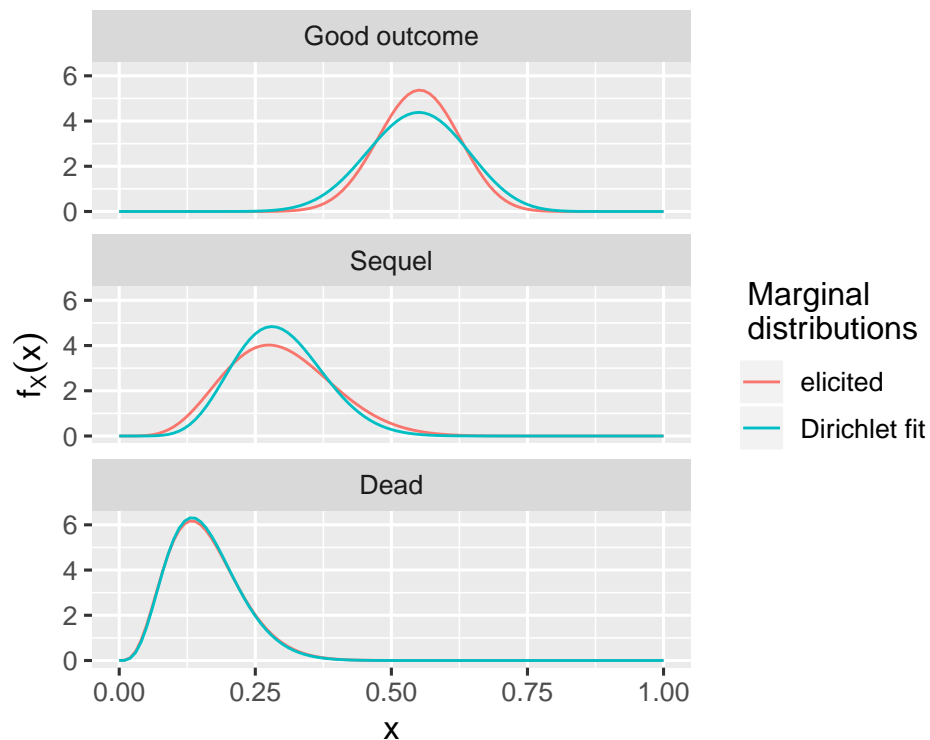
The objects `fit.good`, `fit.seql` and `fit.dead` all include parameters of fitted beta distributions, for example,

```
fit.good$Beta
```

```
##      shape1 shape2
## 1 24.88927 20.4087
```

We can now fit the Dirichlet distribution to the elicited marginals.

```
d.fit <- fitDirichlet(fit.good, fit.seql, fit.dead,
                     categories = c("Good outcome", "Sequel", "Dead"),
                     n.fitted = "opt")
```



```
##
## Directly elicited beta marginal distributions:
##
##      Good outcome  Sequel  Dead
```

```
## shape1      24.9000  6.2000  4.6000
## shape2      20.4000 14.7000 24.4000
## mean        0.5490  0.2960  0.1590
## sd          0.0731  0.0975  0.0667
## sum         45.3000 20.9000 29.0000
##
## Sum of elicited marginal means: 1.004
##
## Beta marginal distributions from Dirichlet fit:
##
##      Good outcome  Sequel    Dead
## shape1          16.6000  8.9600  4.8000
## shape2          13.8000 21.4000 25.6000
## mean            0.5470  0.2950  0.1580
## sd              0.0889  0.0814  0.0651
## sum             30.4000 30.4000 30.4000
```

The above plot shows both the marginal distributions that were elicited directly, and the the marginal distributions resulting from the Dirichlet fit. Parameters and summaries from these two sets of distributions are shown as output. We see that the marginal distribution for the ‘Dead’ proportion hasn’t changed appreciably, but that the Dirichlet fit has resulted in a little more uncertainty for the ‘Good outcome’ proportion, and a little less uncertainty for the ‘Sequel’ proportion.

The Dirichlet parameters are stored in `d.fit`, but can be read off from the `shape1` row: we have  $d_1 = 16.6$ ,  $d_2 = 8.96$ ,  $d_3 = 4.8$  (the values have been rounded for display purposes).

We can report feedback from the marginal distributions of the fitted Dirichlet:

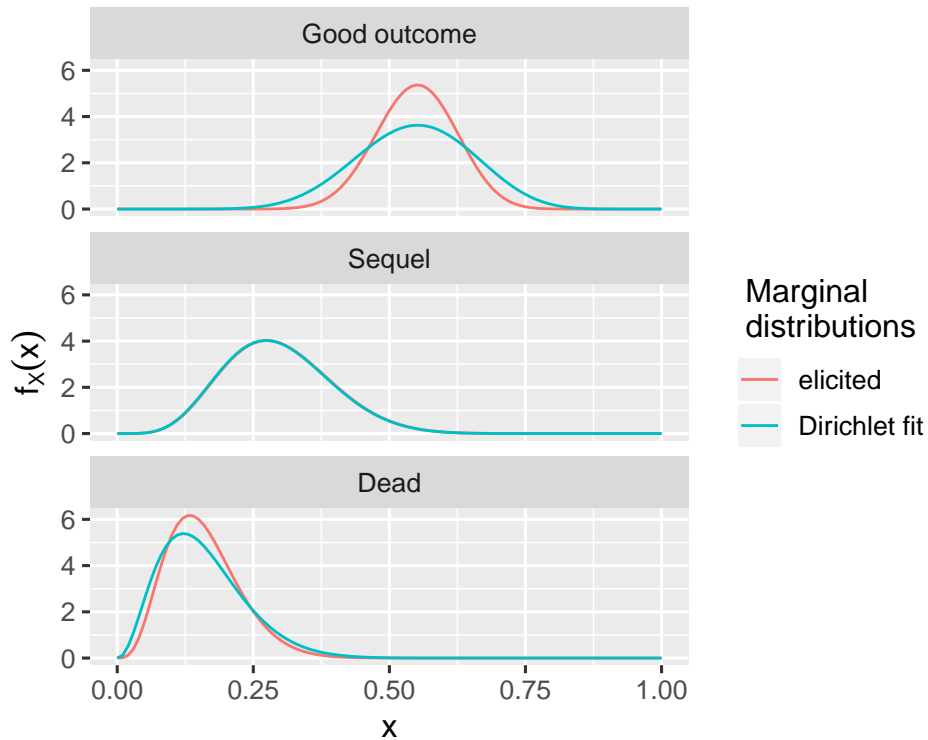
```
feedbackDirichlet(d.fit, quantiles = c(0.1, 0.5, 0.9))
```

```
##      quantiles Good.outcome Sequel Dead
## [1,]      0.1          0.43  0.19 0.08
## [2,]      0.5          0.55  0.29 0.15
## [3,]      0.9          0.66  0.40 0.25
```

so, for example, after fitting the Dirichlet distribution, the fitted median and 90th percentile for the proportion of ‘good outcomes’ are 0.55 and 0.66 respectively.

The parameter  $n$  was chosen by minimising the sum of squared differences between the marginal standard deviations in the elicited marginal beta distributions and the marginals from the fitted Dirichlet. An alternative, more conservative choice is to set  $n$  as the minimum of the sum of the beta parameters in each elicited marginal. From the output above, we can see that this will correspond to the ‘Sequel’ proportion.

```
d.fit <- fitDirichlet(fit.good, fit.seql, fit.dead,
                    categories = c("Good outcome", "Sequel", "Dead"),
                    n.fitted = "min")
```



```
##
## Directly elicited beta marginal distributions:
##
##      Good outcome  Sequel    Dead
## shape1      24.9000  6.2000  4.6000
## shape2      20.4000 14.7000 24.4000
## mean         0.5490  0.2960  0.1590
## sd           0.0731  0.0975  0.0667
## sum          45.3000 20.9000 29.0000
##
## Sum of elicited marginal means: 1.004
##
## Beta marginal distributions from Dirichlet fit:
##
##      Good outcome  Sequel    Dead
## shape1      11.500  6.1800  3.3100
## shape2       9.480 14.8000 17.6000
## mean         0.547  0.2950  0.1580
## sd           0.106  0.0974  0.0779
## sum          20.900 20.9000 20.9000
```

We see that there is almost no change between the elicited and fitted marginal for the ‘Sequel’ proportion, barring a minor adjustment to ensure the fitted marginal means sum to 1.