Dirichlet-Categorical distributions

David Mareček

October 19, 2022

EUROPEAN UNION
European Structural and Investment Fund
Operational Programme Research,
Development and Education

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

Many of the slides in this presentation were taken from the presentations of Carl Edward Rasmussen (University of Cambridge)

Multinomial distribution

Generalisation of the binomial distribution from 2 outcomes to m outcomes.
Useful for random variables that take one of a finite set of possible outcomes.
Throw a die $n = 60$ times, and count the observed (6 possible) outcomes.

Outcome	Count
$X = x_1 = 1$	$k_1 = 12$
$X = x_2 = 2$	$k_2 = 7$
$X = x_3 = 3$	$k_3 = 11$
$X = x_4 = 4$	$k_4 = 8$
$X = x_5 = 5$	$k_5 = 9$
$X = x_6 = 6$	$k_6 = 13$

Note that we have one parameter too many. We don't need to know all the k_i and n, because $\sum_{i=1}^{6} k_i = n$.

Multinomial Distribution

Consider a discrete random variable X that can take one of m values $x_1, \dots, x_m$.
Out of n independent trials, let k_i be the number of times $X = x_i$ was observed.
It follows that $\sum_{i=1}^{m} k_i = n$.
Denote by π_i the probability that $X = x_i$, with $\sum_{i=1}^{m} \pi_i = 1$.
The probability of observing a vector of occurrences $k = [k_1, \dots, k_m]$ is given by the multinomial distribution parametrized by $\pi = [\pi_1, \dots, \pi_m]$.

$$p(k|\pi, n) = p(k_1, \dots, k_m|\pi_1, \dots, \pi_m, n) = \frac{n!}{k_1!k_2!\dots k_m!} \prod_{i=1}^{m} \pi_i^{k_i}$$

- Note that we can write $p(k|\pi)$ since n is redundant.
- The multinomial coefficient $\frac{n!}{k_1!k_2!\dots k_m!}$ is a generalisation of $\binom{n}{k}$.

The *categorical* distribution is the generalisation of the *Bernoulli* distribution to m outcomes, and the special case of the *multinomial* distribution with one trial:

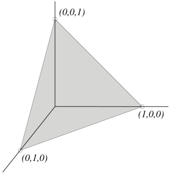
$$p(X = x_i|\pi) = \pi_i$$

Dirichlet Distribution

The *Dirichlet* distribution is to the *Categorical/Multinomial* what the *Beta* distribution is to the *Bernoulli/Binomial*.

It is a generalisation of the Beta defined on the $m-1$ dimensional simplex.

- Consider the vector $\pi = [\pi_1, \ldots, \pi_m]$, with $\sum_{i=1}^m \pi_i = 1$ and $\pi_i \in (0,1)\ \forall i$.
- Vector π lives in the open standard $m-1$ simplex.
- π could for example be the parameter vector of a multinomial.

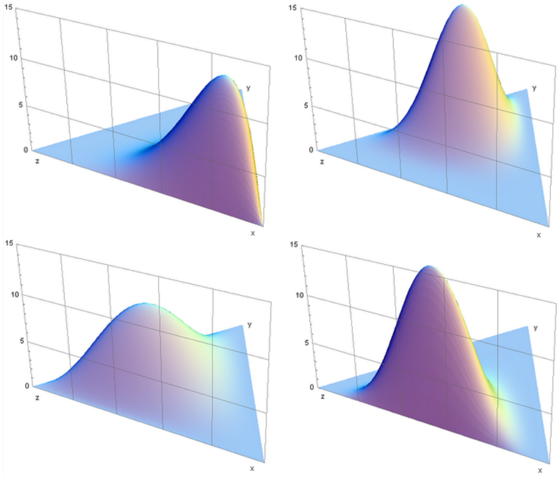


The Dirichlet distribution is given by

$$Dir(\vec{\pi}|\alpha_1, \ldots, \alpha_m) = \frac{\Gamma(\sum_{i=1}^m \alpha_i)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m \pi_i^{\alpha_i - 1} = \frac{1}{B(\alpha)} \prod_{i=1}^m \pi_i^{\alpha_i - 1}$$

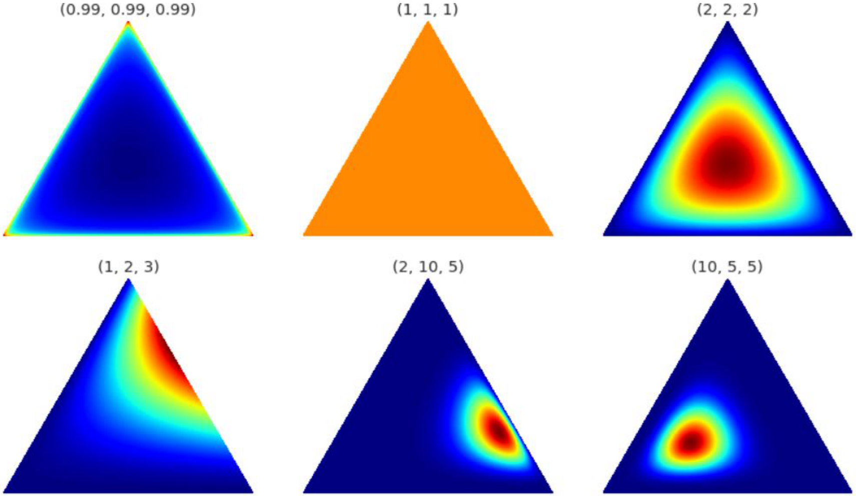
- $\alpha = [\alpha_1, \ldots, \alpha_m]$
- $B(\alpha)$ is the multivariate beta function
- $E(\pi_j) = \frac{\alpha_j}{\sum_i = 1^m \alpha_i}$ is the mean for the j-th element.

Dirichlet Distribution: examples



Several images of the probability density of the Dirichlet distribution when $n = 3$ for various parameter vectors α. Clockwise from top left: $\alpha = (6,2,2),(3,7,5),(6,2,6),(2,3,4)$.

Dirichlet Distribution: examples

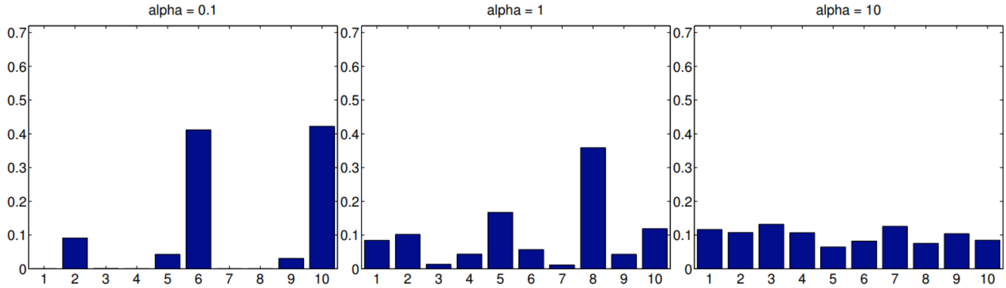


The symmetric Dirichlet distribution

In the symmetric Dirichlet distribution, all parameters are identical: $\alpha_i = \alpha, \ \forall i$.

`en.wikipedia.org/wiki/File:LogDirichletDensity-alpha_0.3_to_alpha_2.0.gif`

Distributions drawn at random from symmetric 10 dimensional Dirichlet distributions with various concentration parameters.



Beta and Dirichlet: definition

Beta distribution:

$$Beta(\pi|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}$$

Dirichlet distribution (generalization of Beta distribution to m outcomes):

$$Dir(\vec{\pi}|\alpha_1,\dots,\alpha_m) = \frac{\Gamma(\sum_{i=1}^{m}\alpha_i)}{\prod_{i=1}^{m}\Gamma(\alpha_i)}\prod_{i=1}^{m}\pi_i^{\alpha_i-1}$$

Beta and Dirichlet: posterior probability

Posterior probability for Beta-Bernoulli distribution:

$$p(\pi|D) = \frac{p(\pi|\alpha,\beta)p(D|\pi)}{p(D)} \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}\pi^{k}(1-\pi)^{n-k} = \pi^{\alpha+k-1}(1-\pi)^{\beta+n-k-1}$$

$$p(\pi|D) = Beta(\pi|\alpha+k,\beta+n-k)$$

Posterior probability for Dirichlet-Categorical distribution:

$$p(\vec{\pi}|D) = \frac{p(\vec{\pi}|\vec{\alpha})p(D|\vec{\pi})}{p(D)} \propto \prod_{i=1}^{m} \pi_i^{\alpha_i-1}\pi_i^{k_i} = \prod_{i=1}^{m} \pi_i^{\alpha_i+k_i-1} \propto Dir(\vec{\pi}|\vec{\alpha}+\vec{k})$$

Beta and Dirichlet: predictive distributions

Beta-Bernoulli predictions:

$$p(x_{next} = 1|D) = \int_0^1 p(x_{next} = 1|\pi)p(\pi|D)d\pi = \int_0^1 \pi Beta(\pi|\alpha{+}k, \beta{+}n{-}k)d\pi = \frac{\alpha + k}{\alpha + \beta + n}$$

Dirichlet-Categorical predictions:

$$p(x_{next} = j|D) = \int_\triangle p(x_{next} = j|\vec{\pi})p(\vec{\pi}|D)d\vec{\pi} = \int_\triangle \pi_j Dir(\vec{\pi}|\vec{\alpha} + \vec{k})d\vec{\pi} = \frac{\alpha_j + k_j}{\sum_{i=1}^m(\alpha_i + k_i)}$$

The sign $\triangle$ indicates a simplex: the integral is taken across all vectors $\vec{\pi}$ that are valid probability distributions, i.e. $\sum_{i=1}^m \pi_i = 1$.

The integrals are equal to expected values of given Beta/Dirichlet posterior distributions.