Dirichlet Distribution

distribution over distribution

Renhao Ye

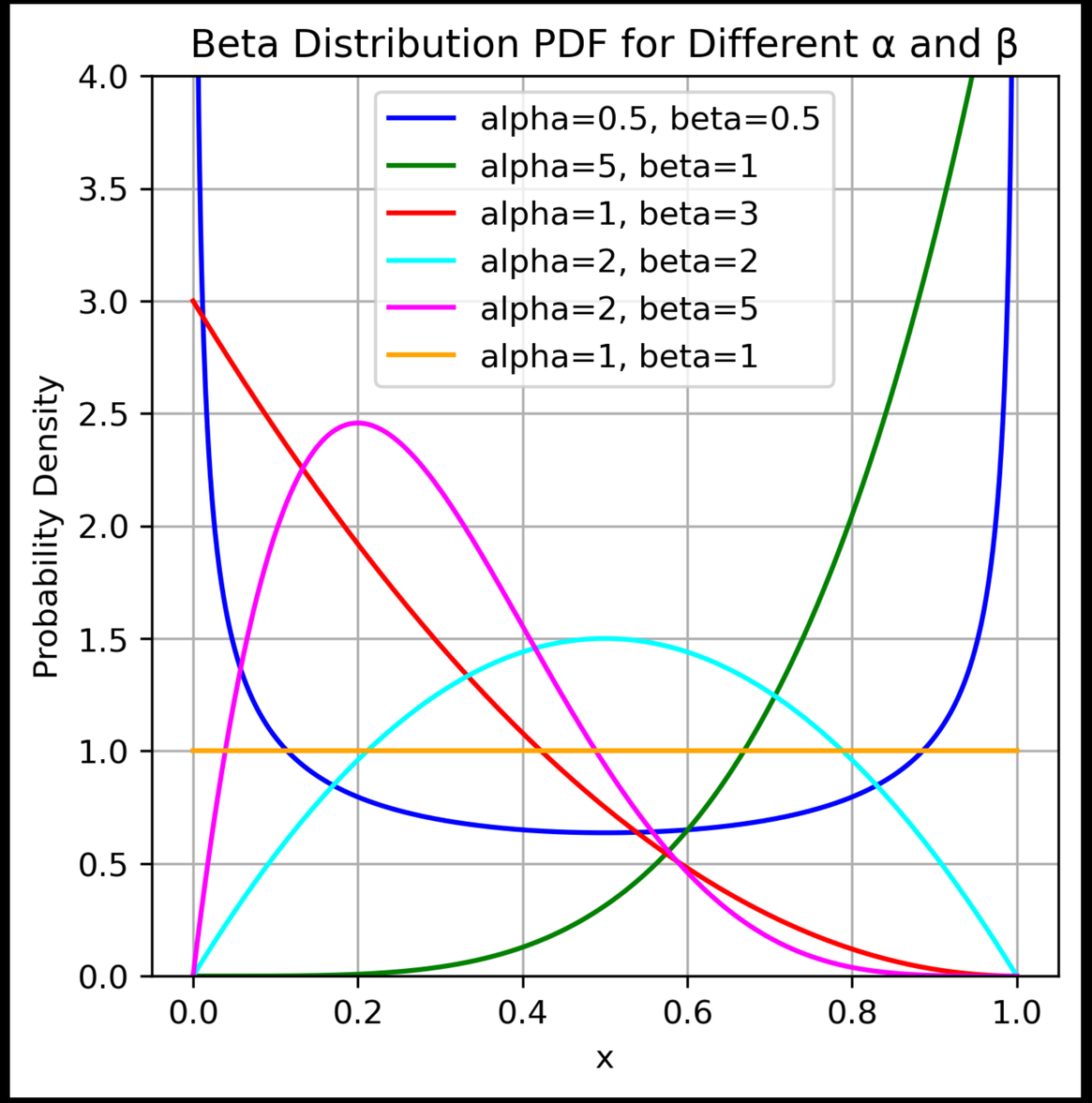
Binomial Distribution

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

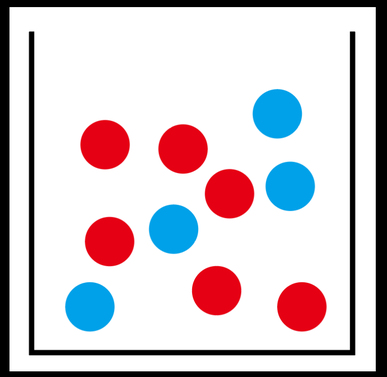
Beta Distribution

Conjunctive prior of Binomial Distribution

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$$

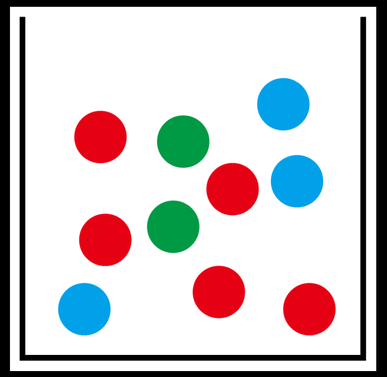


Example



Binomial $N = 10$

Beta $p_{\text{blue}} = 0.4,\ p_{\text{red}} = 0.6$



Multinomial $N = 10$

Dirichlet $p_{\text{blue}} = 0.3,\ p_{\text{red}} = 0.5,\ p_{\text{green}} = 0.2$

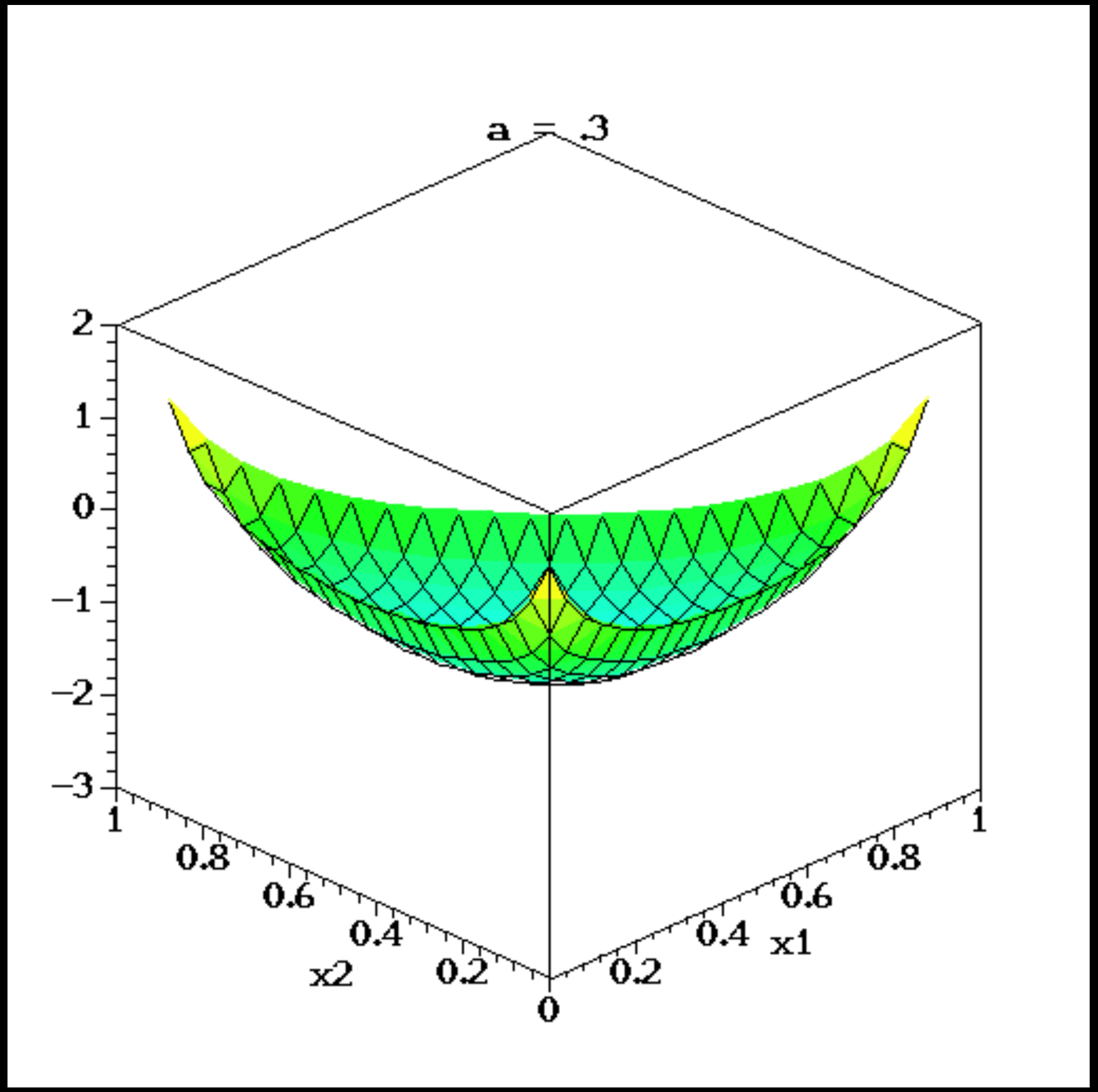
Dirichlet Distribution

$$f(x) = \frac{1}{\mathrm{B}(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

K is the dimension of the distribution

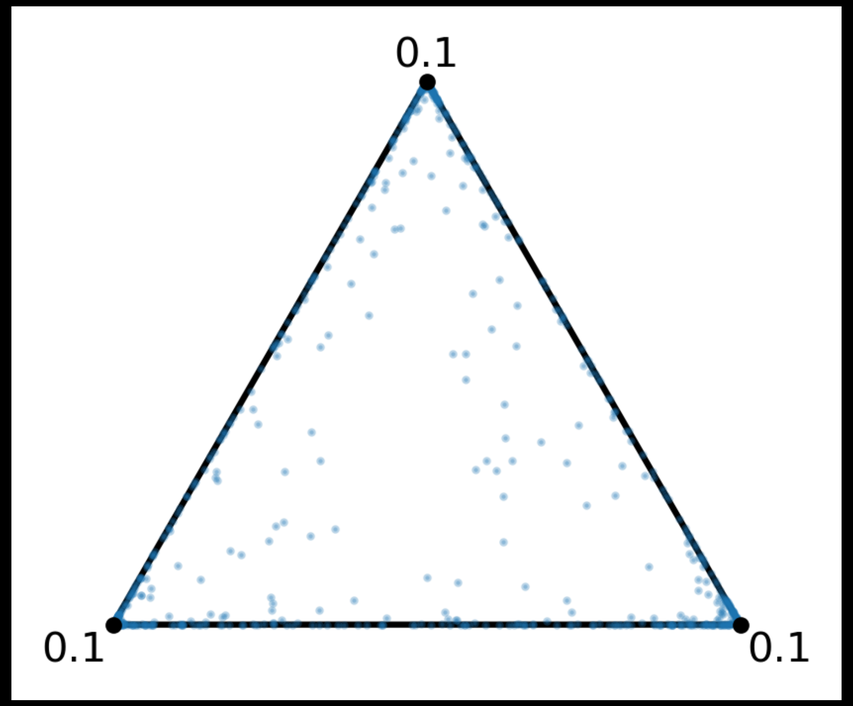
alpha is the concentration parameter

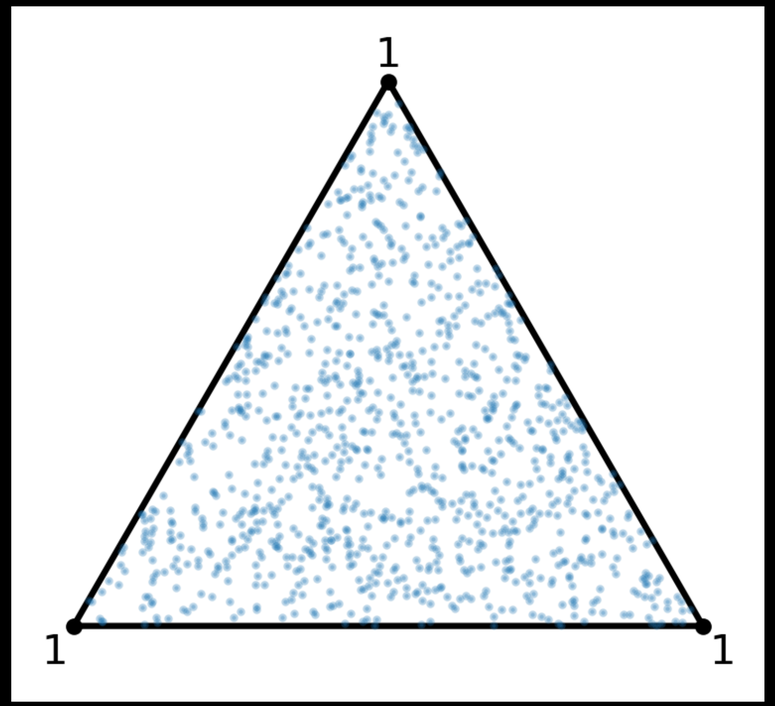
Normalizing factor: $\mathrm{B}(\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\sum_{k=1}^{K} \alpha_i)}$

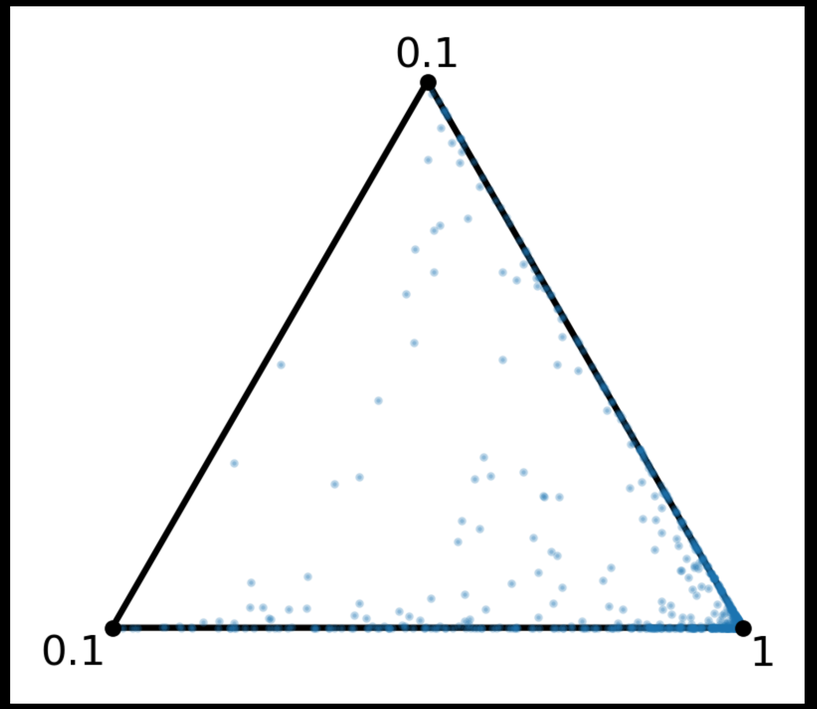

a = .3
2
1
0
−1
−2
−3
1
0.8
0.6
0.4
0.2
0
x2
0.2
0.4
0.6
0.8
1
x1

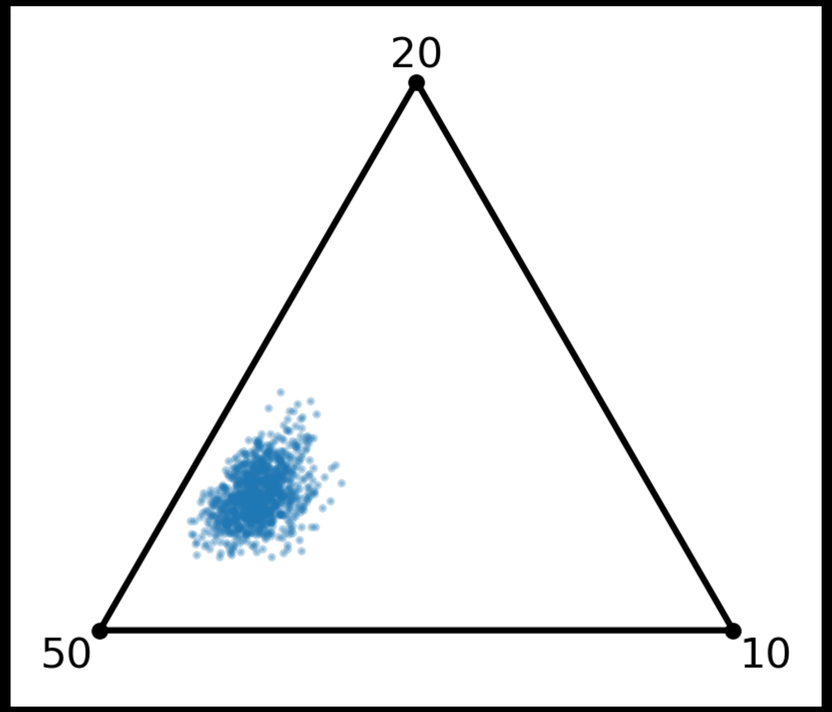
Understanding Concentration

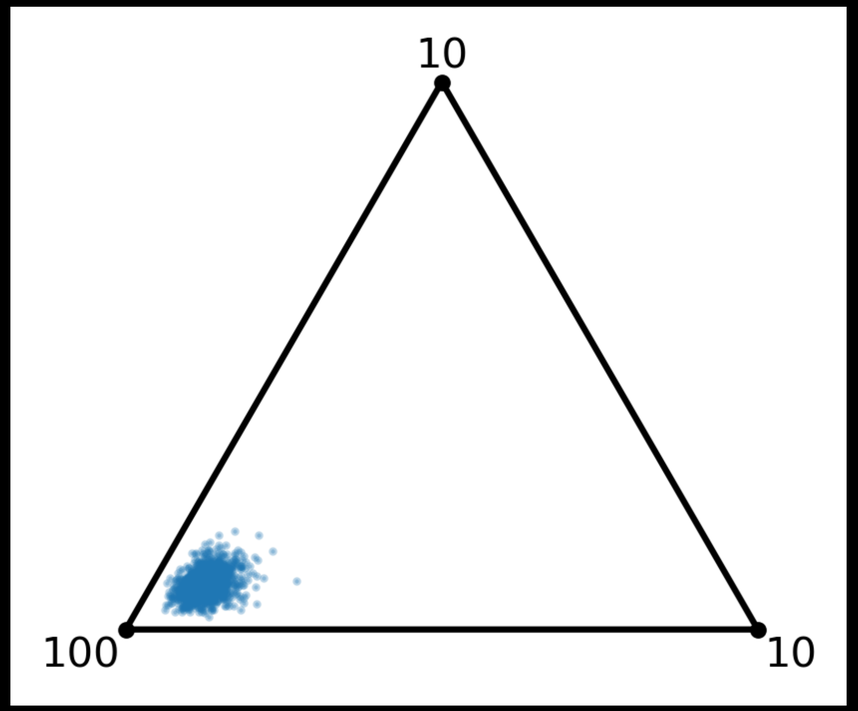
The higher value of α_i, the greater "weight" of X_i and the greater amount of the total "mass" is assigned to it (recall that in total it must be $x_1 + \cdots + x_k = 1$). If all α_i are equal, the distribution is symmetric. If $\alpha_i < 1$, it can be thought of as anti-weight that pushes away x_i toward extremes, while when it is high, it attracts x_i toward some central value (central in the sense that all points are concentrated around it, *not* in the sense that it is symmetrically central). If $\alpha_1 = \cdots = \alpha_k = 1$, then the points are uniformly distributed.

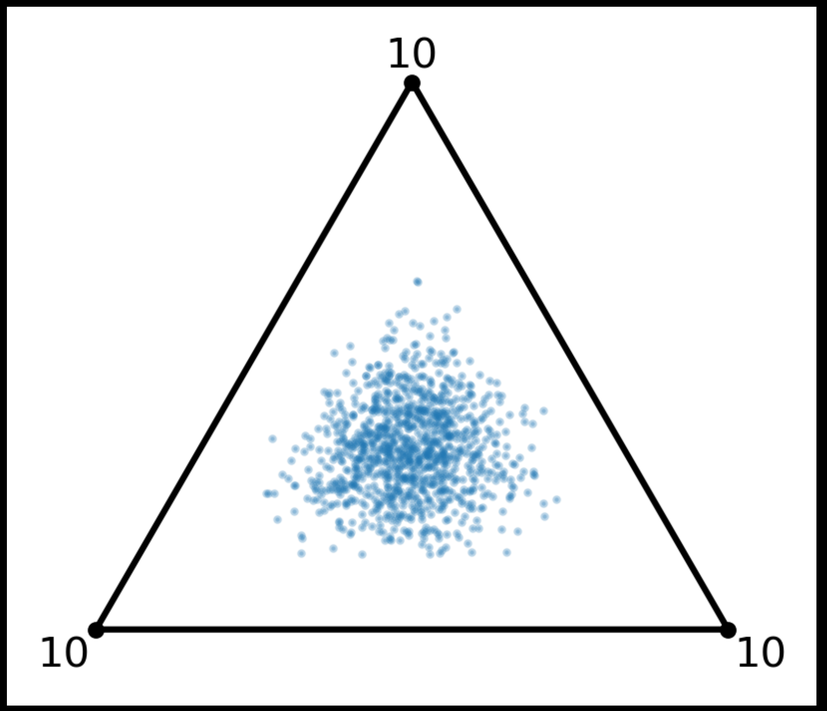

0.1
0.1
0.1


1
1
1


0.1
0.1
1


20
50
10


10
100
10


10
10
10

Reference

- https://stats.stackexchange.com/questions/244917/what-exactly-is-the-alpha-in-the-dirichlet-distribution/
- https://zhuanlan.zhihu.com/p/425388698
- https://en.wikipedia.org/wiki/Dirichlet_distribution