Probability Distributions: Continuous

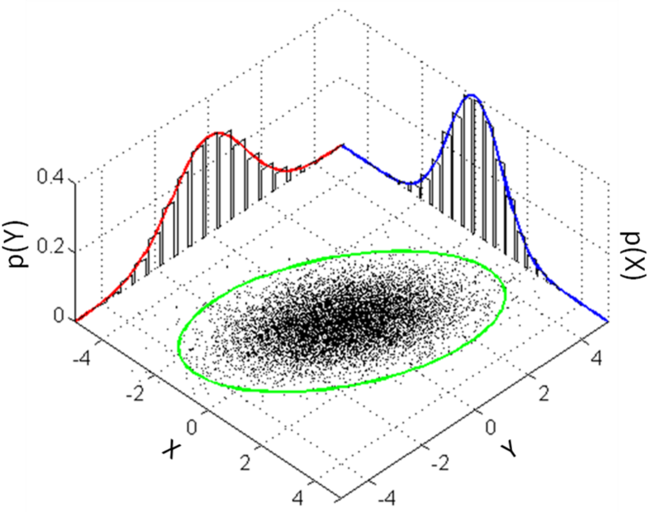
Introduction to Data Science Algorithms

Dirk Grunwald

Multivariate normal distribution

- What is the *joint* distribution over multiple normal variables?
- If the normal random variables are independent, the joint distribution is just the product of each individual PDF.
- But they don't have to be independent.
- We can model the joint distribution over multiple variables with the *multivariate* normal distribution.

Multivariate normal distribution



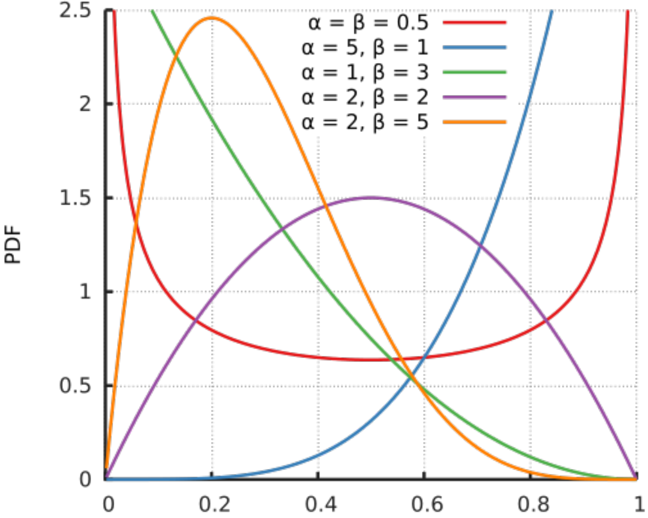
Multivariate normal distribution

- The multivariate normal distribution is a distribution over a *vector* of values $\mathbf{x}$. The mean μ is also a vector.
- In addition to the variance of each variable, each pair of variables has a *covariance*.
  - The covariance matrix for all pairs is denoted Σ.
  - The covariance indicates an association between variables. If it is positive, it means if one value increases (or decreases), the other value is also likely to increase (or decrease). If the covariance is negative, it means that if one value increases, the other is likely to decrease, and vice versa.
- $f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right)$

Beta distribution

- The Beta distribution is over real values on the interval $[0, 1]$.
- Useful distribution for modeling percentages and proportions
  - Batting averages in baseball
  - Percentage of people with a disease in a country
- The density is proportional to: $x^{\alpha-1}(1-x)^{\beta-1}$
- Related to the Bernoulli distribution: $x^{\theta}(1-x)^{1-\theta}$

Beta distribution



Beta distribution

- The PDF for Beta is:

$$f(x) = \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}}_{\substack{\text{Inverse Beta function,}\\ \text{related to the binomial coefficient}}} x^{\alpha-1}(1-x)^{\beta-1}$$

- Γ is the gamma function, $\Gamma(x) = (x-1)!$
  - Just like the factorial function, but works for real values in addition to integers
- Mean: $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$
- The parameters α and β must be > 0.

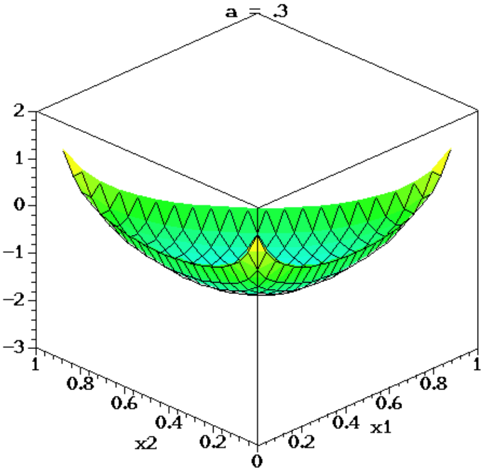
Dirichlet distribution

- The Dirichlet distribution is a generalization of the Beta distribution for multiple random variables
- The Dirichlet distribution is over *vectors* whose values are all in the interval $[0, 1]$ and the sum of values in the vector is 1.
  - In other words, the vectors in the sample space of the Dirichlet have the same properties as probability distributions.
  - The Dirichlet distribution can be thought of as a “distribution over distributions”.

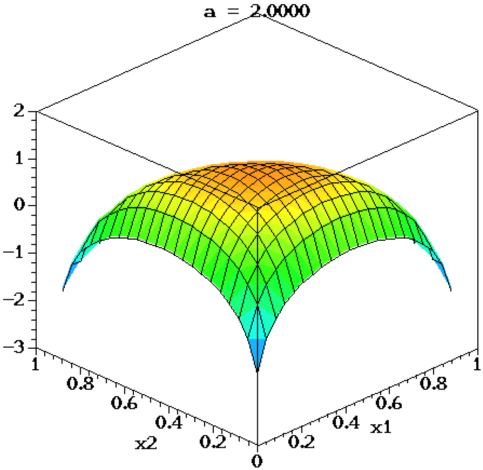
- The PDF for a K-dimensional Dirichlet distribution has a vector of parameters denoted α, given by:

$$f(\mathbf{x}) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_i\right)} \prod_{k=1}^{K} x_k^{\alpha_k - 1}$$

Dirichlet distribution



Dirichlet distribution



Dirichlet distribution

- The Dirichlet PDF looks similar to the multinomial distribution.
  - The Dirichlet density is proportional to: $\prod_k x_k^{\alpha_k - 1}$
  - The multinomial mass is proportional to: $\prod_k x_k^{\theta_k}$

- Remember this analogy:
  - **Beta : binomial :: Dirichlet : multinomial**