Dirichlet Distribution

Sushil Vemuri

About Me!

- Bachelors of Science (B.S) from Indian Institute of Science Education and Research, Bhopal
- Currently in Fourth Year
- Major: Data Science and Engineering Department
- Minor: Electrical Engineering and Computer Science Department
- Interests: Perception and Motion Planning in Autonomous Vehicles, NLP

CONTENTS

01

Binomial Distribution

Binomial Distribution

Discrete probability Distribution of the number of successes in a sequence of independent experiments, with a known probability of success for the experiment.

- Random Variable:
  X = Number of Successes

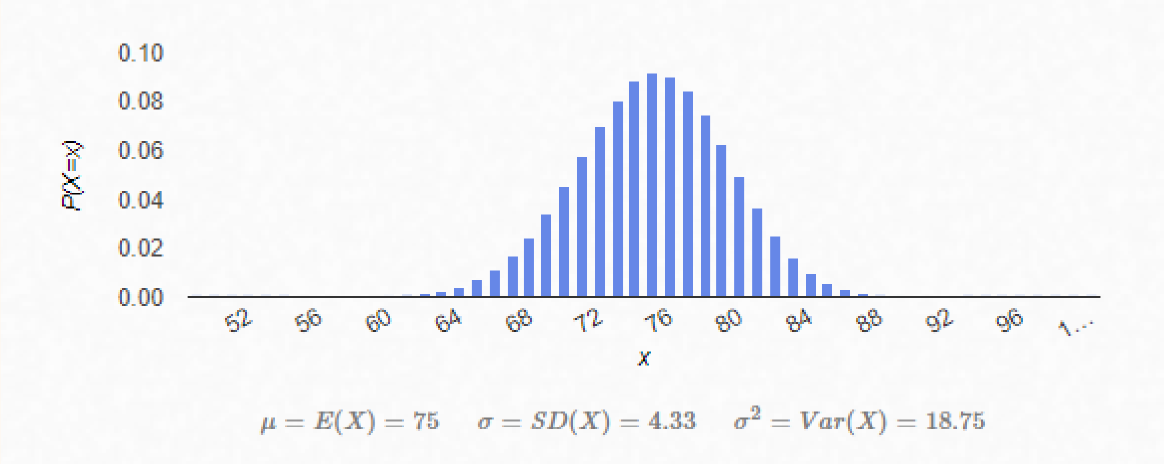
- Parameters:
  n = Number of experiments
  p = Probability of success of an experiment

- Probability Mass Function

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

⇒ Probability of x successes in n experiments where the probability of success of each experiment is p.

Binomial Distribution



$n = 100$

$p = 0.75$

02

Beta Distribution

Beta Distribution

Continuous probability Distribution over the probability of Success of a single experiment in a Binomial Distribution with a known amount of Successes and Failures.

- Random Variable:
  X = Probability of Success of a single experiment

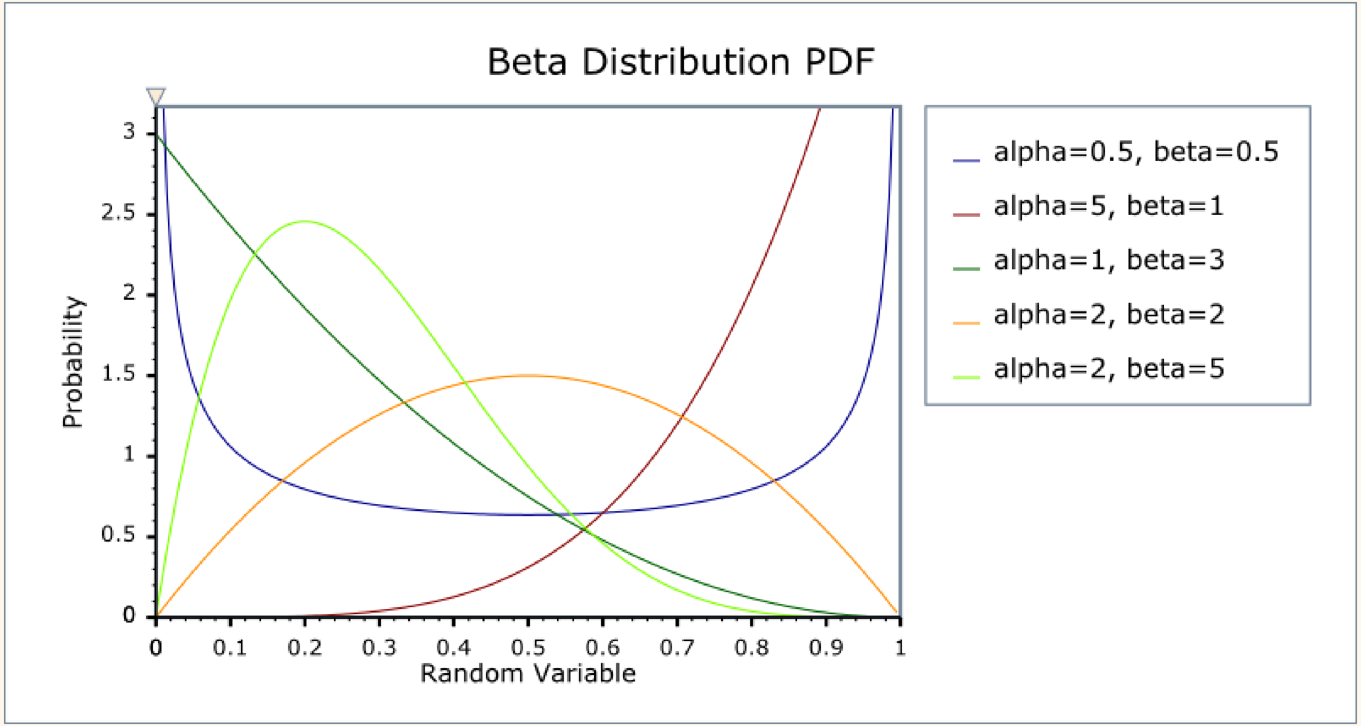
- Parameters:
  α = Number of Successes + 1
  β = Number of Failures + 1

- Probability Density Function

$$\mathrm{f}(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

⇒ Probability of probability of success being x in a binomial distribution where there were $\alpha - 1$ successful experiments and $\beta - 1$ failed experiments.

Beta Distribution



03

Multinomial Distribution

Multinomial Distribution

Discrete probability Distribution of the number of successes of each class in a sequence of independent experiments, with a know probability of success for each class. It is a multivariate generalization of binomial distribution.

- Random Variable:
  $X = [X_1, X_2, \dots, X_k]$, where X_i is the number of successes of class i.

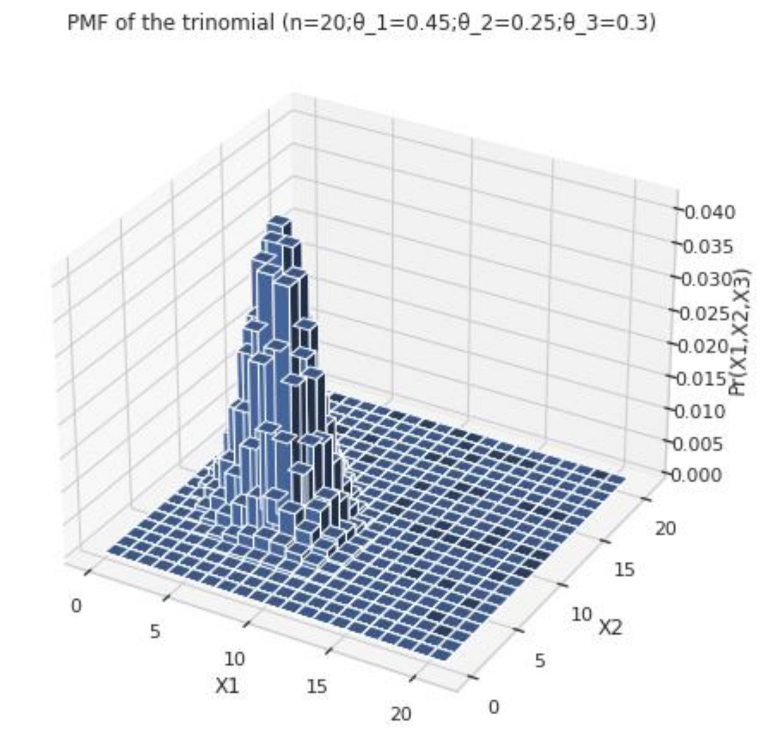
- Parameters:
  n = Number of experiments
  k = Number of classes
  $p = [p_1, p_2, \dots, p_k]$, where p_i is the probability of success of class i.

- Probability Mass Function

$$P(X = x) = \begin{cases} \dfrac{n!}{x_1!\, x_2! \dots x_k!}\, p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} & \text{if } \sum_{i=1}^{k} x_i = n \\ 0 & \text{otherwise} \end{cases}$$

  ⇒ Probability of x_i successes of class i with probability of success p_i for all i, in n experiments for k classes.

Multinomial Distribution



04

Dirichlet Distribution

Dirichlet Distribution

Continuous probability Distribution over the probability of Success of every class in a Multinomial Distribution with a known amount of Successes per class. It is a multivariate generalization of the beta distribution.

- Random Variable:
  $X = [X_1, X_2, \ldots, X_k]$, where X_i is the Probability of Success of class i.

- Parameters:
  $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_k]$, where $\alpha_i - 1$ is the number of successes of class i.

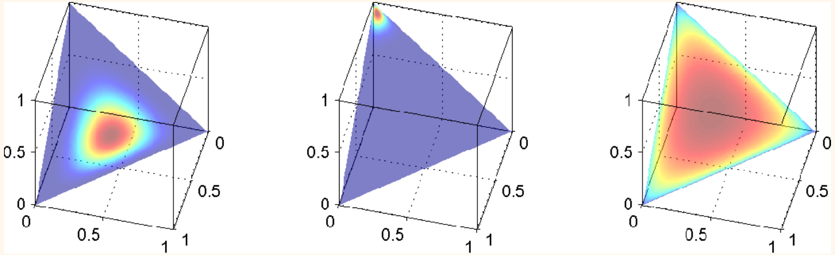
- Probability Density Function

$$\mathrm{f}(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} \prod_{i=1}^{k} x_i^{\alpha_i - 1} & \text{if } 0 \leq x_i \leq 1 \text{ and } \sum_{i=1}^{k} x_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

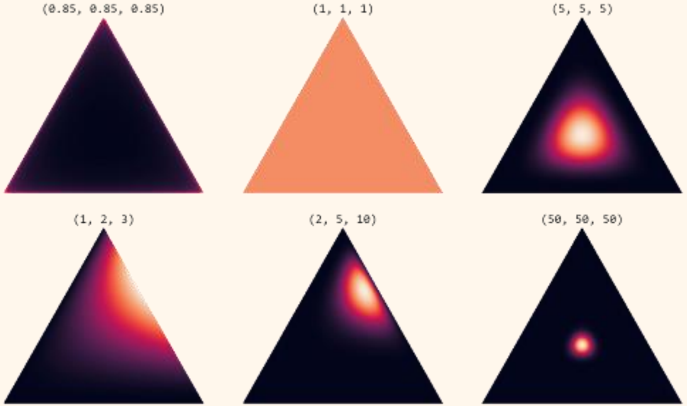
⇒ Probability of probability of success being x_i for class i in a multinomial distribution where there were class i had α_i successful experiments.

Dirichlet Distribution

The PDF has a value only when $0 \leq x_i \leq\ 1$ and $\sum_{i=1}^{k} x_i = 1$. When k=3, this is a triangle in the first quadrant of the 3 graph.



Dirichlet Distribution for 3 classes.



Dirichlet Distribution

The Dirichlet distribution is the conjugate prior for the categorical distribution and multinomial distributions in Bayesian inference.

For some likelihood functions, if you choose a certain prior, the posterior ends up being in the same distribution as the prior. Such a prior then is called a Conjugate Prior.

$$\overset{\text{posterior}}{P(\theta|X)} = \frac{\overset{\text{sampling}}{P(X|\theta)} \cdot \overset{\text{prior}}{P(\theta)}}{\int_{\underset{\text{for every possible } \theta}{\theta}} P(X|\theta) \cdot P(\theta)\, d\theta} \rightarrow \text{normalizing constant}$$

Computing a posterior using a conjugate prior is very convenient, because you can avoid expensive numerical computation involved in Bayesian Inference.

Dirichlet Distribution

Formally this can be expressed as follows,
Given a model

$$\begin{array}{rcccl} \boldsymbol{\alpha} & = & (\alpha_1, \ldots, \alpha_K) & = & \text{concentration hyperparameter} \\ \mathbf{p} \mid \boldsymbol{\alpha} & = & (p_1, \ldots, p_K) & \sim & \operatorname{Dir}(K, \boldsymbol{\alpha}) \\ \mathbb{X} \mid \mathbf{p} & = & (\mathbf{x}_1, \ldots, \mathbf{x}_K) & \sim & \operatorname{Cat}(K, \mathbf{p}) \end{array}$$

Then the following holds

$$\begin{array}{rcccl} \mathbf{c} & = & (c_1, \ldots, c_K) & = & \text{number of occurrences of category } i \\ \mathbf{p} \mid \mathbb{X}, \boldsymbol{\alpha} & \sim & \operatorname{Dir}(K, \mathbf{c} + \boldsymbol{\alpha}) & = & \operatorname{Dir}(K, c_1 + \alpha_1, \ldots, c_K + \alpha_K) \end{array}$$

Dirichlet Distribution

Let $X = (X_1, \ldots, X_K) \sim \mathrm{Dir}(\alpha)$. Then,

$$\mathrm{E}[X_i] = \frac{\alpha_i}{\alpha_0},$$

$$\mathrm{Var}[X_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}.$$

$$\mathrm{Cov}[X_i, X_j] = \frac{-\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}.$$

The mode of the distribution is[7] the vector (x_1, ..., x_K) with

$$x_i = \frac{\alpha_i - 1}{\alpha_0 - K}, \qquad \alpha_i > 1.$$

THANK YOU!