Pattern Recognition

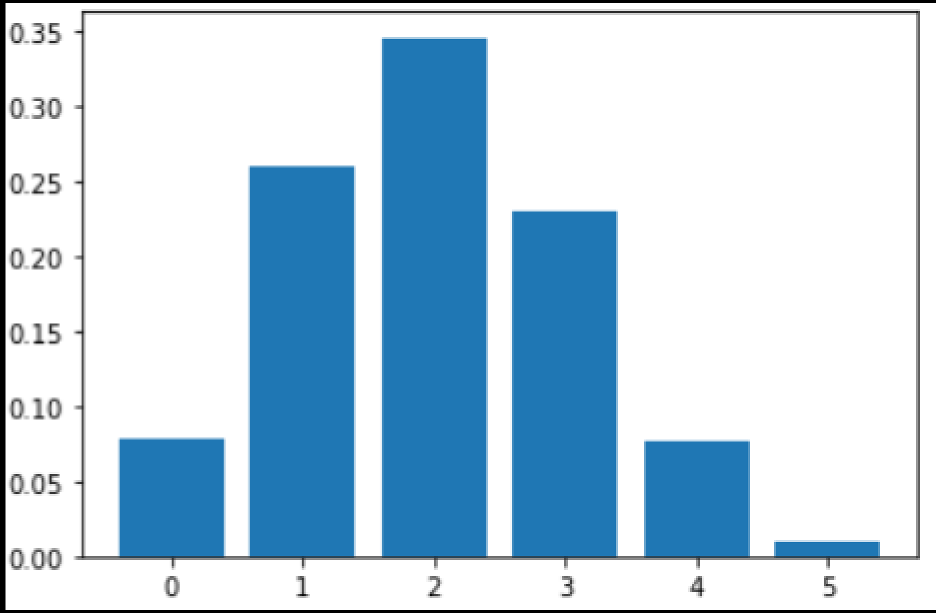
Probability Distributions Revision

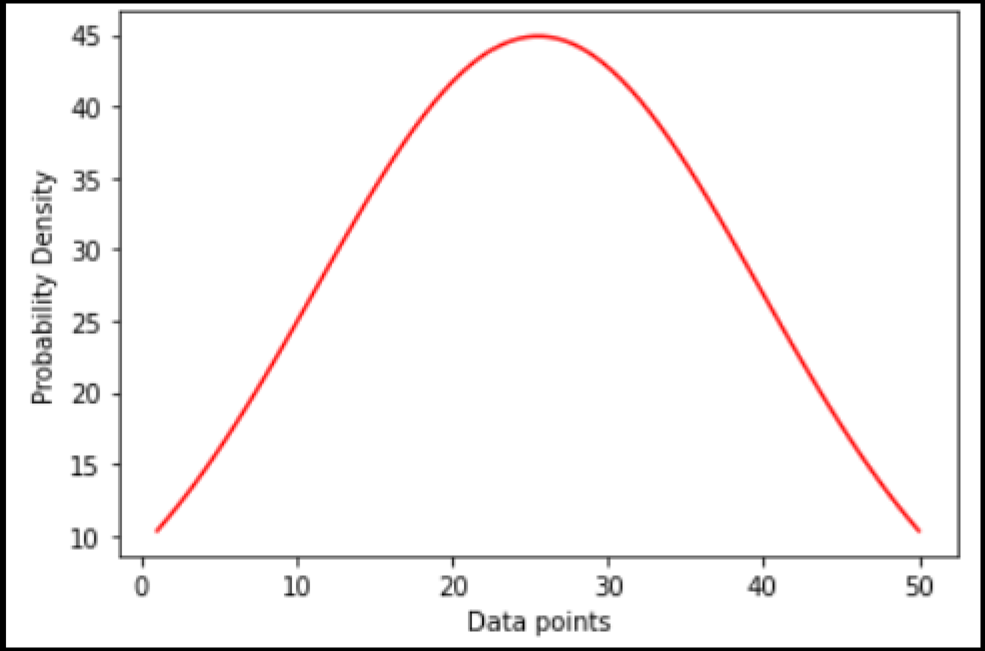
Content

Content
Probability Distributions
Bernoulli Distribution
Multinomial Distribution

Probability Distributions

- A probability distribution gives the possibility of each outcome of a random experiment or event.
  - Probability is a measure of uncertainty of various phenomena.



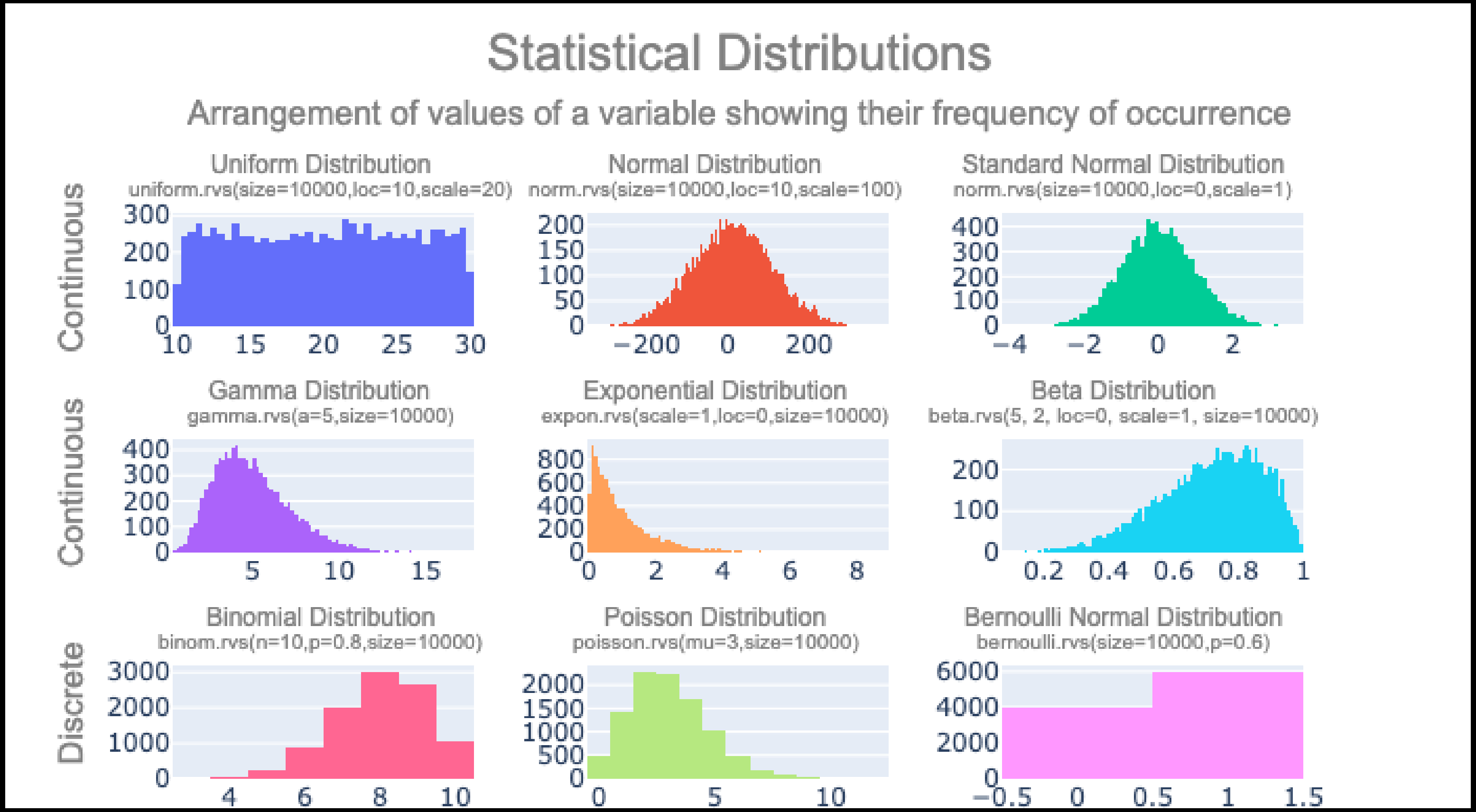


Probability Distributions

- Probabilities help us make predictions.
- Given a probabilistic output, we can always compute our “best guess” as to the “true label”
  - The most probable class label
- Google has a system known as SmartASS (ad selection system) that predicts the probability you will click on an ad based on your search history and other user and ad-specific features. This probability is known as the click-through rate or CTR and can be used to maximize expected profit.

Probability Distributions

- You can find common probability distributions and their applications at https://en.wikipedia.org/wiki/Probability_distribution



https://miro.medium.com/max/700/1*cr9_-ts4vqVBOttf-EVuQQ.png

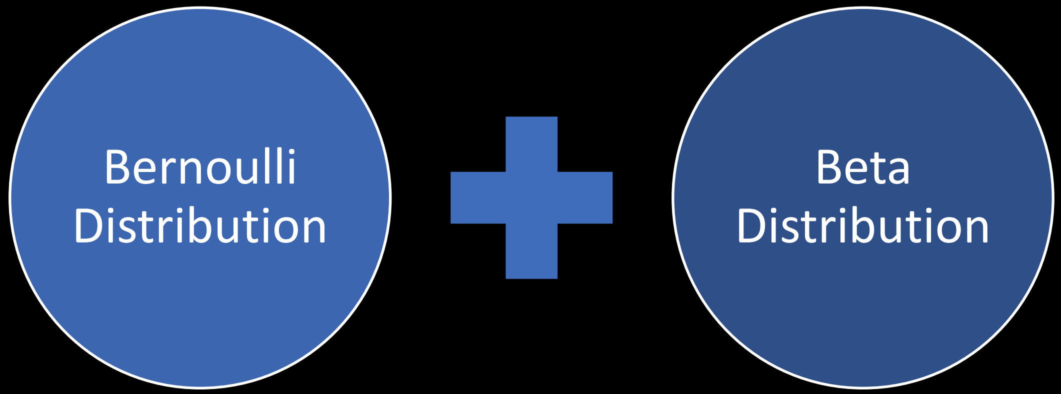
Probability Distributions

- We explored the following distributions

Bernoulli Distribution

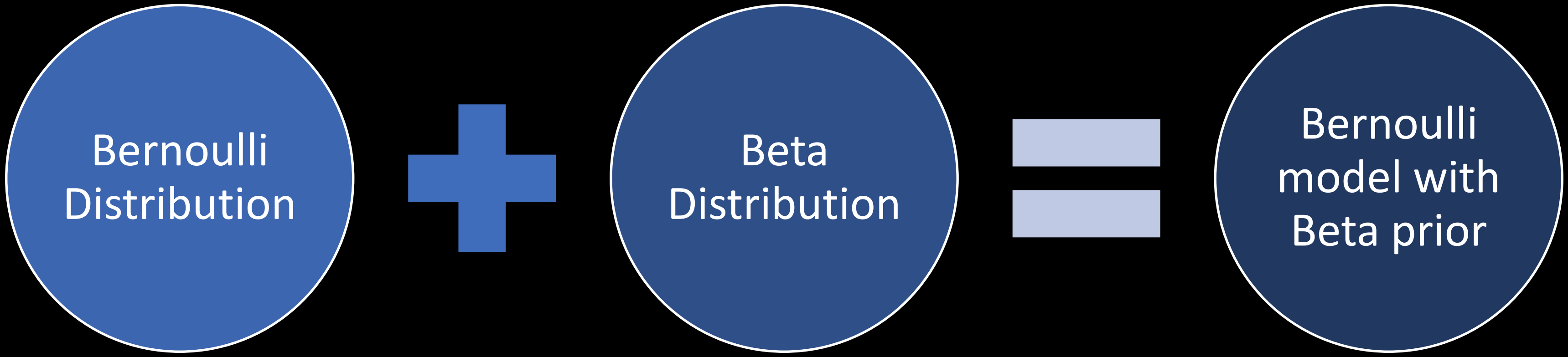
Probability Distributions

- We explored the following distributions



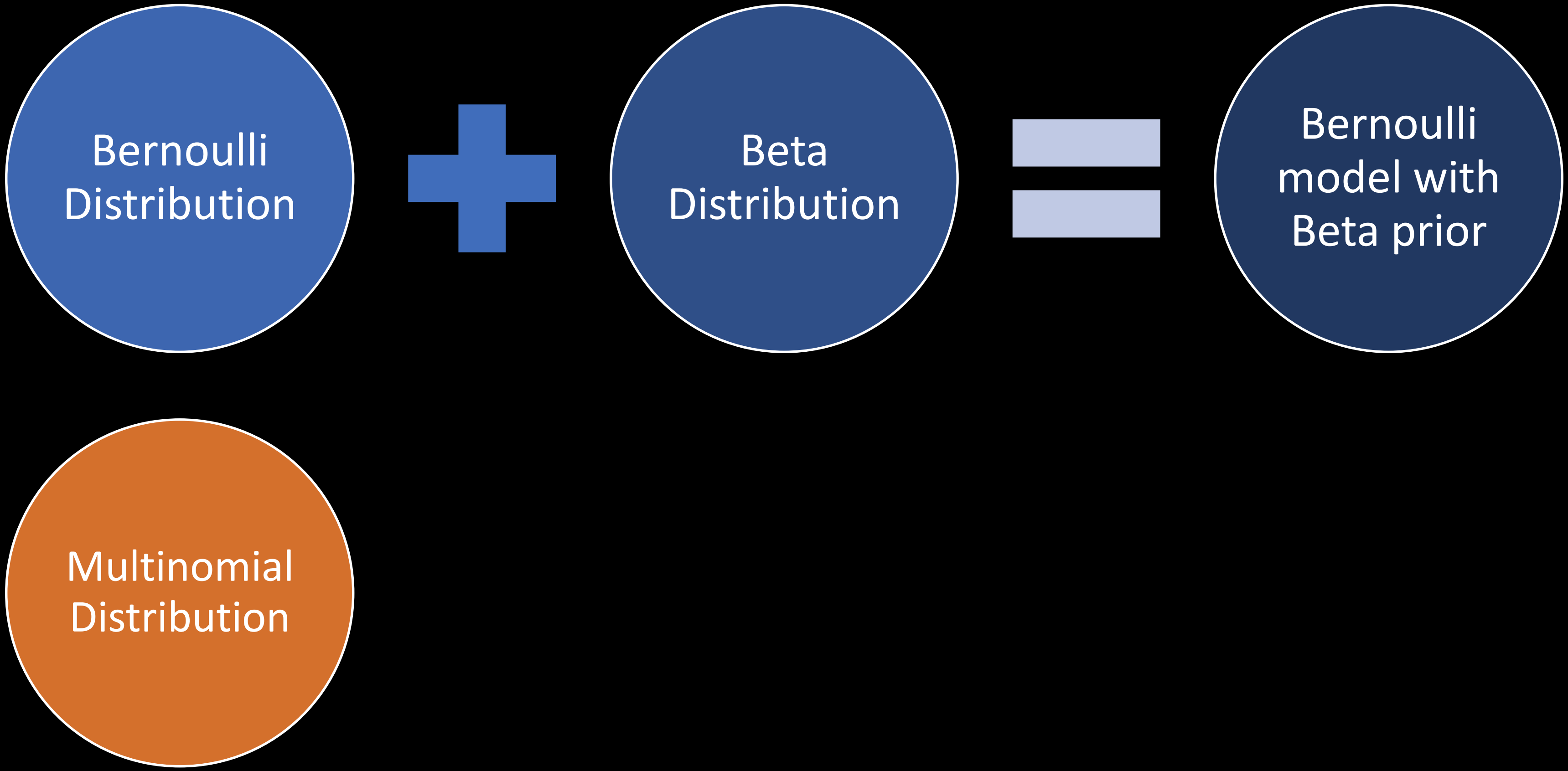
Probability Distributions

- We explored the following distributions



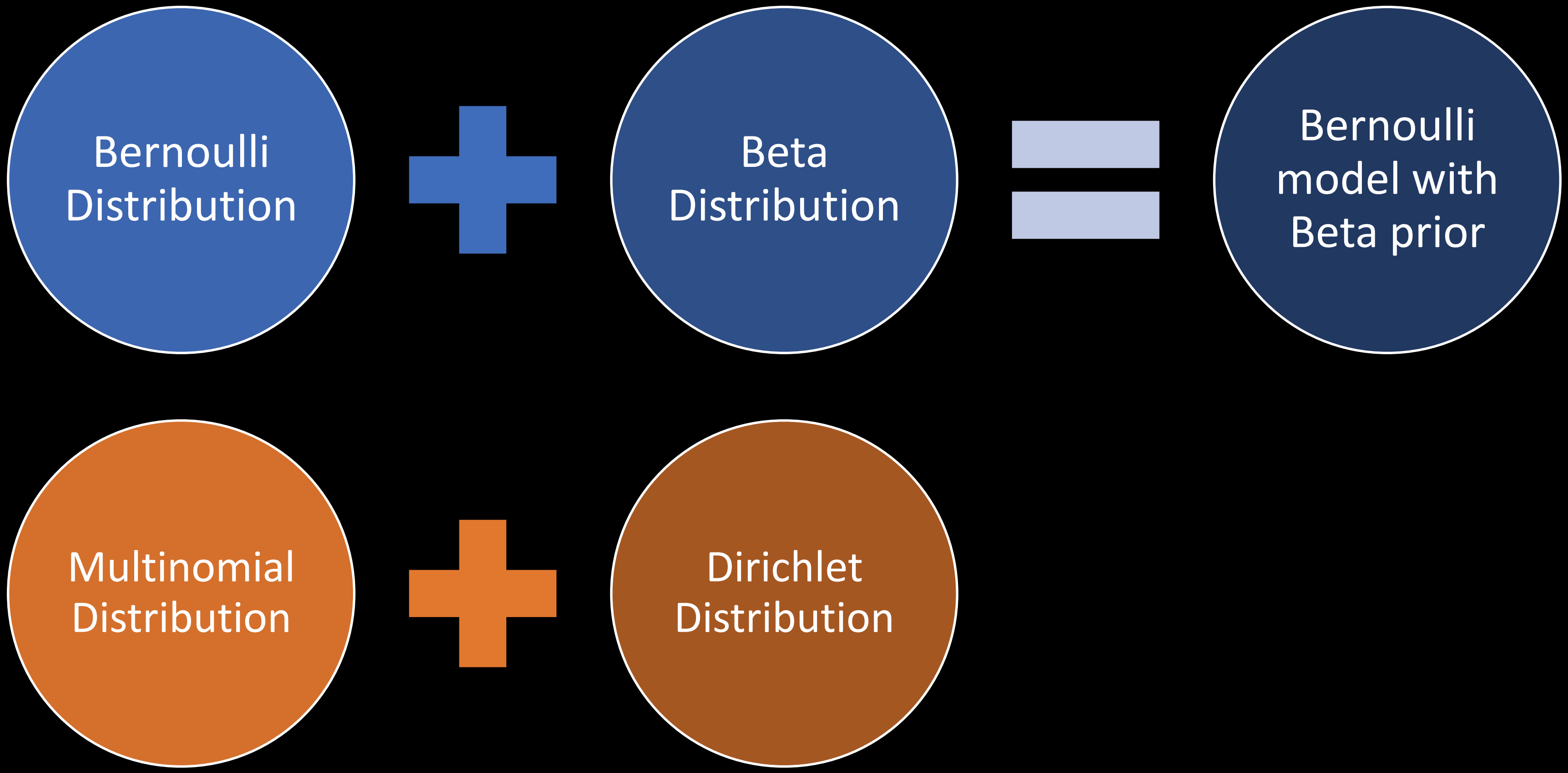
Probability Distributions

- We explored the following distributions



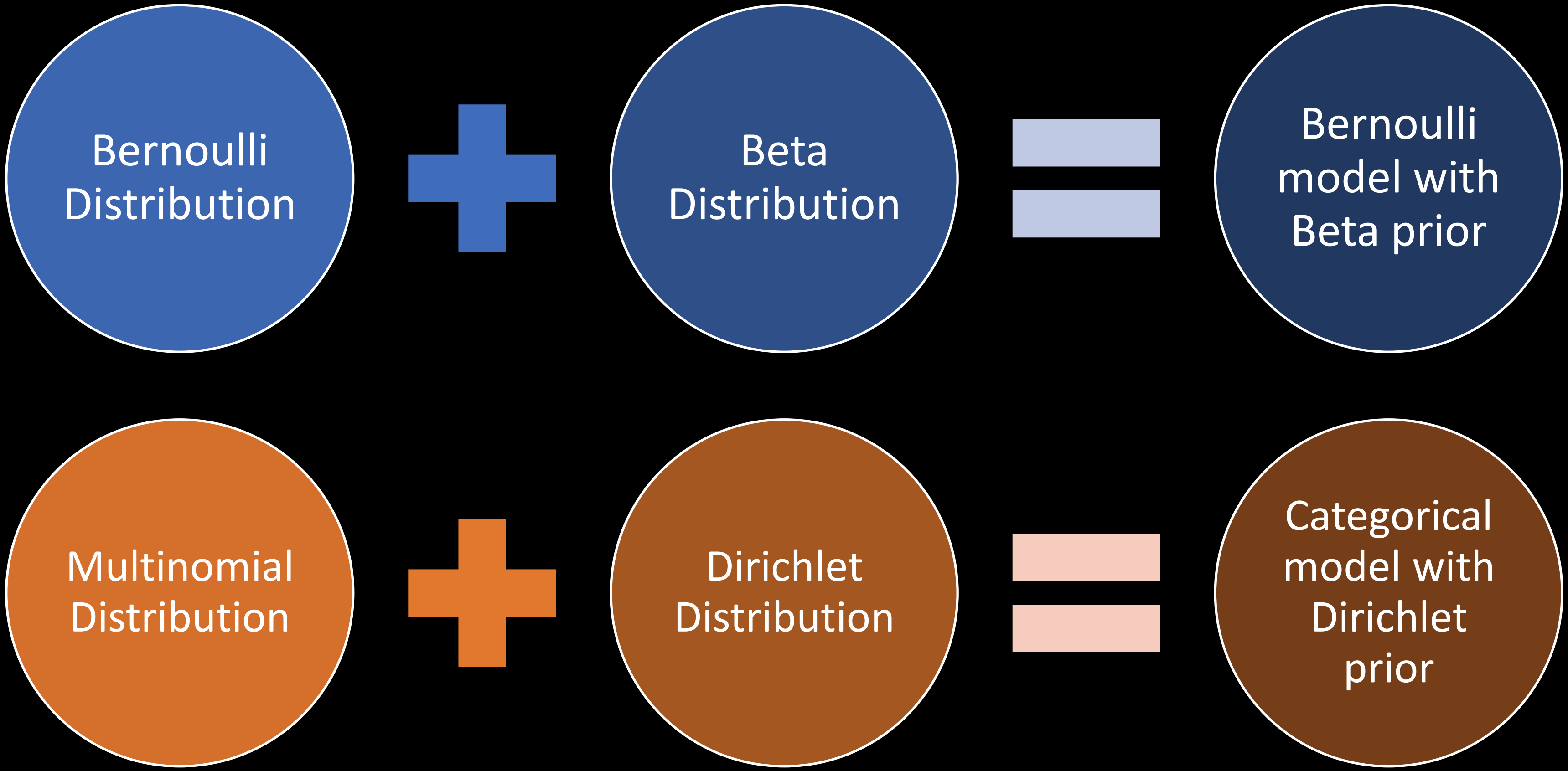
Probability Distributions

- We explored the following distributions



Probability Distributions

- We explored the following distributions



Content

Content
Probability Distributions
Bernoulli Distribution
Multinomial Distribution

Bernoulli Distribution

- Bernoulli distribution is a discrete probability distribution where variables can have value 0 or 1.

$$Bern(x|\mu) = \mu^x (1-\mu)^{1-x}$$

  - x is the random variable
  - μ is the probability that $x = 1$
- The output depends on μ.

Bernoulli Distribution

There are two ways to compute μ:

1. Maximum likelihood estimation: $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$

2. Using Beta distribution as a prior distribution:

$$Beta(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}$$

- $\Gamma(x)$ is a gamma function.
- a is the number of objects with x=1.
- b is the number of objects with x=0.

Bernoulli Distribution

- To use Beta as a prior for computing μ in the Bernoulli distribution, we get the model as

$$p(x = 1|D) = \frac{m + a}{m + a + l + b}$$

  - m is the number of objects = 1
  - $l =$ number of elements in the dataset - m

Bernoulli Distribution

- Example 1: using maximum likelihood estimation

Email	Spam?
email 1	1
email 2	1
email 3	1
email 4	1
email 5	1
email 6	0
email 7	0
email 8	0
email 9	0
email 10	1

$$Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$$

1. Compute μ:
$$\frac{1+1+1+1+1+0+0+0+0+1}{10} = 0.6$$
2. $P(x = 1|\mu = 0.6) = 0.6^1(1 - 0.6)^{1-1} = 0.6$
3. $P(x = 0|\mu = 0.6) = 0.6^0(1 - 0.6)^{1-0} = 0.4$

Bernoulli Distribution

- Example 2: using maximum likelihood estimation

Email	Spam?
email 1	1
email 2	1
email 3	1
email 4	1
email 5	1
email 6	1
email 7	1
email 8	1
email 9	1
email 10	1

$$Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$$

1. Compute μ:
$$\frac{1+1+1+1+1+1+1+1+1+1}{10} = 1$$
2. $P(x = 1|\mu = 1) = 1^1(1-1)^{1-1} = 1$
3. $P(x = 0|\mu = 1) = 1^0(1-1)^{1-0} = 0$

Bernoulli Distribution

- Example 3: using Beta prior

Email	Spam?
email 1	1
email 2	1
email 3	1
email 4	1
email 5	1
email 6	1
email 7	1
email 8	1
email 9	1
email 10	1

$$Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$$
$$p(x=1|D) = \frac{m+a}{m+a+l+b}$$

1. Compute μ (suppose $a = 2, b = 3$):
$$\frac{10+2}{10+2+0+3} = 0.8$$
2. $P(x = 1|\mu = 0.8) = 0.8^1(1-0.8)^{1-1} = 0.8$
3. $P(x = 0|\mu = 0.8) = 0.8^0(1-0.8)^{1-0} = 0.2$

Content

Content
Probability Distributions
Bernoulli Distribution
Multinomial Distribution

Multinomial Distribution

- Given the vector $\boldsymbol{x}$ and the vector $\boldsymbol{\mu}$, the distribution of $\boldsymbol{x}$ is given by

$$p(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

  - $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots \mu_k)^T$ and the parameter μ_k is the probability that x=1.
- The output depends on μ

Multinomial Distribution

There are two ways to compute μ

1. Using maximum likelihood estimation

$$\mu_k^{ML} = \frac{m_k}{N}$$

2. Using Dirichlet distribution as a prior

$$Dir(\boldsymbol{\mu}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1), \ldots \Gamma(\alpha_K)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}$$

- $\boldsymbol{\alpha}$ are the parameters that control the distributions.
- $\alpha_0 = \alpha_1 + \alpha_2 + \cdots + \alpha_k$

Multinomial Distribution

To use Dirichlet distribution as a prior inside multinomial distribution, we get the new $\boldsymbol{\alpha}$ parameters as follows

$$\alpha`_k = \alpha_k + m_k$$

Where the new $\boldsymbol{\alpha}$s can be normalized and used in the multinomial model.

Multinomial Distribution

Example 4: using maximum likelihood estimation

$$p(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

Day	Cloudy	Rainy	Sunny
day 1	1	0	0
day 2	0	1	0
day 3	1	0	0
day 4	0	0	1
day 5	0	1	0
day 6	1	0	0
day 7	0	0	1
day 8	0	0	1
day 9	0	1	0
Day 10	0	0	1

1. Compute μ_1:
$$\frac{1+0+1+0+0+1+0+0+0+0}{10} = 0.3$$
2. Compute μ_2:
$$\frac{0+1+0+0+1+0+0+0+1+0}{10} = 0.3$$
3. Compute μ_3:
$$\frac{0+0+0+1+0+0+1+1+0+1}{10} = 0.4$$

1. $P(cloudy|[0.3, 0.3, 0.4]) = 0.3^1 * 0.3^0 * 0.4^0 = 0.3$
2. $P(rainy|[0.3, 0.3, 0.4]) = 0.3^0 * 0.3^1 * 0.4^0 = 0.3$
3. $P(sunny|[0.3, 0.3, 0.4]) = 0.3^0 * 0.3^0 * 0.4^1 = 0.4$

Multinomial Distribution

$$p(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

Example 5: using maximum likelihood estimation

Day	Cloudy	Rainy	Sunny
day 1	0	0	1
day 2	0	0	1
day 3	0	0	1
day 4	0	0	1
day 5	0	0	1
day 6	0	0	1
day 7	0	0	1
day 8	0	0	1
day 9	0	0	1
Day 10	0	0	1

1. Compute μ_1:
$$\frac{0+0+0+0+0+0+0+0+0+0}{10} = 0$$
2. Compute μ_2:
$$\frac{0+0+0+0+0+0+0+0+0+0}{10} = 0$$
3. Compute μ_3:
$$\frac{1+1+1+1+1+1+1+1+1+1}{10} = 1$$

1. $P(cloudy|[0,0,1]) = 0^1 * 0^0 * 1^0 = 0$
2. $P(rainy|[0,0,1]) = 0^0 * 0^1 * 1^0 = 0$
3. $P(sunny|[0,0,1]) = 0^0 * 0^0 * 1^1 = 1$

Multinomial Distribution

Example 6: using Dirichlet distribution as a prior

Day	Cloudy	Rainy	Sunny
day 1	0	0	1
day 2	0	0	1
day 3	0	0	1
day 4	0	0	1
day 5	0	0	1
day 6	0	0	1
day 7	0	0	1
day 8	0	0	1
day 9	0	0	1
Day 10	0	0	1

1. $\mu_1 = 0$
2. $\mu_2 = 0$
3. $\mu_3 = 1$

$$p(\boldsymbol{x}|\boldsymbol{\mu}) = \prod_{k=1}^{K} \mu_k^{x_k}$$

$$\alpha`_k = \alpha_k + m_k$$

1. Suppose $\alpha_1 = 2, \alpha_2 = 3, \alpha_3 = 5.$
2. we have $m_1 = 0, m_2 = 0, m_3 = 10$
3. $\alpha`_1 = 2 + 0 = 2$
4. $\alpha`_2 = 3 + 0 = 3$
5. $\alpha`_3 = 5 + 10 = 15$

Multinomial Distribution

https://courses.cs.duke.edu/fall14/compsci527/notes/Dirichlet.pdf

https://www.cs.ubc.ca/~nando/340-2012/lectures/l19.pdf

https://stephentu.github.io/writeups/dirichlet-conjugate-prior.pdf

https://leimao.github.io/blog/Introduction-to-Dirichlet-Distribution/#Binomial-Distribution

Machine learning – a Probabilistic Perspective