

COMP2610 / COMP6261 - Information Theory

Lecture 5: Bernoulli, Binomial, Maximum Likelihood and MAP

Assignment Project Exam Help

<https://eduassistpro.github.io>



Australian
National
University

Add WeChat edu_assist_pro

6 August 2018

Assignment Project Exam Help

- Exa

▶ <https://eduassistpro.github.io>

- Frequentist vs Bayesian probabilities

Add WeChat edu_assist_pr

The Bayesian Inference Framework

Bayesian Inference

Bayesian inference provides us with a mathematical framework explaining how to change our (prior) beliefs in the light of new evidence

likelihood prior

— — . .

<https://eduassistpro.github.io>

evidence

$p(X|Z)$

$= \frac{p(X) p(Z|X)}{p(Z)}$

Add WeChat: edu_assist_pro

Prior: Belief that someone is sick

Likelihood: Probability of testing positive given someone is sick

Posterior: Probability of being sick given someone tests positive

This time

Assignment Project Exam Help

- The Bernoulli and binomial distribution (we will make much use of this henceforth in studying binary channels)

<https://eduassistpro.github.io>

- Esti

Add WeChat edu_assist_pr

- Bayesian inference for parameter estimati

Outline

1 The Bernoulli Distribution

2 The Bi

3 Para

4 Bayesian Parameter Estimation

5 Wrapping up

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

- 1 The Bernoulli Distribution

Assignment Project Exam Help

- 2 The Binomial Distribution

- 3 Para <https://eduassistpro.github.io>

- 4 Bayesian Parameter Estimation

Add WeChat edu_assist_pr

- 5 Wrapping up

The Bernoulli Distribution

Introduction

Consider a binary variable $X \in \{0, 1\}$. It could represent many things:

- Whether a coin lands heads or tails

- The

- A tra

- The success of a medical trial

Often, these outcomes (0 or 1) are not equally likely

What is a general way to model such an X ?

The Bernoulli Distribution

Definition

Assignment Project Exam Help

The variable X takes on the outcomes

<https://eduassistpro.github.io>

Here, $0 \leq \theta \leq 1$ is a **parameter** representing the

For higher values of θ , it is more likely to see 1 than 0

- e.g. a biased coin

The Bernoulli Distribution

Definition

By definition,

Assignment Project Exam Help

$$p(X = 1|\theta) = \theta$$

More suc

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

The Bernoulli Distribution

Definition

By definition,

$$p(X = 1|\theta) = \theta$$

More suc

<https://eduassistpro.github.io>

This is known as a Bernoulli distribution over binary o

Add WeChat edu_assist_pro

$$p(X = x|\theta) = \text{Bern}(x|\theta) =$$

Note the use of the conditioning symbol for θ ; will revisit later

The Bernoulli Distribution

Mean and Variance

The **expected value** (or mean) is given by:

$$\mathbb{E}[X|\theta] = \sum_{x \in \{0,1\}} x \cdot p(x|\theta)$$

<https://eduassistpro.github.io>

The **variance** (or squared standard deviation) is given by

$$\begin{aligned}\mathbb{V}[X|\theta] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[(X - \theta)^2] \\ &= (0 - \theta)^2 \cdot p(X = 0|\theta) + (1 - \theta)^2 \cdot p(X = 1|\theta) \\ &= \theta(1 - \theta).\end{aligned}$$

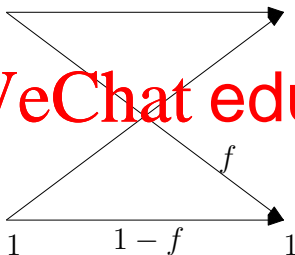
Example: Binary Symmetric Channel

Suppose a sender transmits messages s that are sequences of bits

The receiver sees the bit sequence (message)

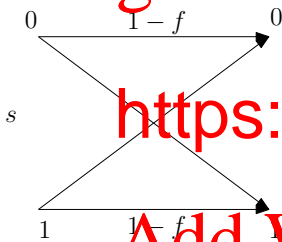
Due to noi

$$0 \leq f \leq$$



Example: Binary Symmetric Channel

We can think of r as the outcome of a random variable, with conditional distribution given by:



If E denotes whether an error occurred, clearly

$$p(E = e) = \text{Bern}(e|f), \quad e \in \{0, 1\}.$$

1 The Bernoulli Distribution

2 The Binomial Distribution

3 Para

<https://eduassistpro.github.io>

4 Bayesian Parameter Estimation

Add WeChat edu_assist_pr

5 Wrapping up

The Binomial Distribution

Introduction

Suppose we perform N independent Bernoulli trials

- e.g. we toss a coin N times

- e.g.

<https://eduassistpro.github.io>

Each trial has

θ

What is the distribution of the number of times (

- e.g. the number of times we obtained

- e.g. the number of errors in the transmitted sequence

The Binomial Distribution

Definition

Let

$Y = \sum_{i=1}^N X_i$

Assignment Project Exam Help

where X_i

Then Y

<https://eduassistpro.github.io>

$$p(Y = m) = \text{Bin}(m|N, \theta) =$$

Add WeChat edu_assist_pro

for $m \in \{0, 1, \dots, N\}$. Here

$$\binom{N}{m} = \frac{N!}{(N-m)!m!}$$

is the # of ways we can we obtain m heads out of N coin flips

The Binomial Distribution:

Mean and Variance

It is easy to show that:

$$\mathbb{E}[Y] = \sum_{m=0}^N m \cdot \text{Bin}(m|N, \theta) = N\theta$$

<https://eduassistpro.github.io>

- Follows from linearity of mean and variance

Add WeChat edu_assist_pro

$$\mathbb{E}[Y] = \mathbb{E}\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N \mathbb{E}[X_i]$$

$$\mathbb{V}[Y] = \mathbb{V}\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N \mathbb{V}[X_i] = N\theta(1 - \theta)$$

The Binomial Distribution:

Example

Ashton is an excellent off spinner. The probability of him getting a wicket during a cricket match is $\frac{1}{4}$. (That is, on each attempt, there is a $1/4$ chance he will get a wicket.)

His coach
game.

- 1 What is the probability that he will get 3 wickets?
 $\text{Bin}(3|10, 0.25)$
- 2 What is the expected number of wickets he will get?
 $\mathbb{E}[Y]$, where $Y \sim \text{Bin}(\cdot|10, 0.25)$.
- 3 What is the probability that he will get at least one wicket?
 $\sum_{m=1}^{10} \text{Bin}(m|N = 10, \theta = 0.25) = 1 - \text{Bin}(m = 0|N = 10, \theta = 0.25)$

The Binomial Distribution:

Example: Distribution of the Number of Wickets

Assignment Project Exam Help

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Histogram of the binomial distribution with $N = 10$ and $\theta = 0.25$. From Bishop (PRML, 2006)

1 The Bernoulli Distribution

Assignment Project Exam Help

2 The Binomial Distribution

3 Para <https://eduassistpro.github.io>

4 Bayesian Parameter Estimation

Add WeChat edu_assist_pr

5 Wrapping up

The Bernoulli Distribution: Parameter Estimation

Consider the set of observations $\mathcal{D} = \{x_1, \dots, x_n\}$ with $x_i \in \{0, 1\}$

- The outcomes of a sequence of coin flips

- Wh

<https://eduassistpro.github.io>

Each observation is the outcome of a random variable

Add WeChat edu_assist_pro

$$p(X = x) = \text{Bern}(x|\theta) =$$

for some parameter θ

The Bernoulli Distribution: Parameter Estimation

We know that

Assignment Project Exam Help

$$X \sim \text{Bern}(x|\theta) = \theta^x(1 - \theta)^{1-x}$$

But often,

- The
- The probability of the word defence a sports

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

What would be a reasonable estimate for θ from \mathcal{D} ?

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

Assignment Project Exam Help

Say that w

<https://eduassistpro.github.io>

Intuitively, which seems more plausible: θ

Add WeChat edu_assist_pro

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

Say that we observe

$$\mathcal{D} = \{0, 0, 0, 1, 0, 0, 1, 0, 0, 0\}$$

If it were tr

<https://eduassistpro.github.io>

$$p(\mathcal{D}|\theta) = \prod_{i=1}^n p(x_i|\theta)$$

Add WeChat: edu_assist_pro

$$\begin{aligned} &= \prod_{i=1}^{10} \frac{1}{2} \\ &= \frac{1}{2^{10}} \\ &\approx 0.001. \end{aligned}$$

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

Say that we observe

$$\mathcal{D} = \{0, 0, 0, 1, 0, 0, 1, 0, 0, 0\}$$

If it were true

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

$$\begin{aligned} p(\mathcal{D}|\theta) &= \prod_{i=1}^{10} p(x_i) \\ &= \left(\frac{1}{5}\right)^2 \cdot \bar{5} \\ &\approx 0.007. \end{aligned}$$

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

We can write down how likely \mathcal{D} is under the Bernoulli model. Assuming independent observations.

<https://eduassistpro.github.io>

We call $L(\theta) = p(\mathcal{D}|\theta)$ the **likelihood** function

Add WeChat edu_assist_pr

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

We can write down how likely \mathcal{D} is under the Bernoulli model. Assuming independent observations.

<https://eduassistpro.github.io>

We call $L(\theta) = p(\mathcal{D}|\theta)$ the **likelihood** function

Add WeChat edu_assist_pro

The parameter for which the observed sequence has the highest probability

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

Maximising $p(\mathcal{D}|\theta)$ is equivalent to maximising $\mathcal{L}(\theta) = \log p(\mathcal{D}|\theta)$

Assignment Project Exam Help

$$\mathcal{L}(\theta) = \log \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i}$$

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

Maximising $p(\mathcal{D}|\theta)$ is equivalent to maximising $\mathcal{L}(\theta) = \log p(\mathcal{D}|\theta)$

Assignment Project Exam Help

$$\mathcal{L}(\theta) = \log \prod_{i=1}^N \theta^{y_i} (1 - \theta)^{1 - y_i}$$

<https://eduassistpro.github.io>

Setting $\frac{d\mathcal{L}}{d\theta} = 0$ we obtain:

Add WeChat edu_assist_pr

$$\theta_{\text{ML}} = \frac{1}{N} \sum_{i=1}^N y_i$$

The Bernoulli Distribution: Parameter Estimation:

Maximum Likelihood

Maximising $p(\mathcal{D}|\theta)$ is equivalent to maximising $\mathcal{L}(\theta) = \log p(\mathcal{D}|\theta)$

$$\mathcal{L}(\theta) = \log \prod_{i=1}^N \theta^{x_i} (1-\theta)^{1-x_i} = \sum_{i=1}^N [x_i \log \theta + (1-x_i) \log(1-\theta)]$$

<https://eduassistpro.github.io>

Setting $\frac{d\mathcal{L}}{d\theta} = 0$ we obtain:

$$\theta_{\text{ML}} = \frac{1}{N} \sum_{i=1}^N x_i$$

The proportion of times $x = 1$ in the dataset \mathcal{D} !

The Bernoulli Distribution:

Parameter Estimation — Issues with Maximum Likelihood

Consider the following scenarios:

- After $N = 3$ coin flips we obtained 3 'tails'
 - ▶ What is the estimate of the probability of a coin flip resulting in 'heads'?
- In a scenario where the probability of success is never 0 or 1, the MLE estimate never exists.

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

The Bernoulli Distribution:

Parameter Estimation — Issues with Maximum Likelihood

Consider the following scenarios:

- After $N = 3$ coin flips we obtained 3 ‘tails’
 - ▶ What is the estimate of the probability of a coin flip resulting in ‘heads’?
- In a scenario where the probability of success is never observed
- ▶ <https://eduassistpro.github.io>

These issues are usually referred to as **overfitting**

- Need to “smooth” out our parameter estimates
- Alternatively, we can do Bayesian inference by considering **priors** over the parameters

1 The Bernoulli Distribution

Assignment Project Exam Help

2 The Binomial Distribution

3 Para

<https://eduassistpro.github.io>

4 Bayesian Parameter Estimation

Add WeChat edu_assist_pr

5 Wrapping up

The Bernoulli Distribution:

Parameter Estimation: Bayesian Inference

Recall:

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)}$$

The equation is annotated with colored brackets: a blue bracket above $p(X|\theta)$ is labeled "likelihood", and a green bracket above $p(\theta)$ is labeled "prior".

If we treat θ as a random variable, we can talk about its value

- e.g. we believe θ is probably close to 0

Our **prior** on θ quantifies what we believe θ is likely to be, **before** looking at the data

Our **posterior** on θ quantifies what we believe θ is likely to be, **after** looking at the data

The Bernoulli Distribution:

Parameter Estimation: Bayesian Inference

The likelihood of X given θ is

Assignment Project Exam Help

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

For the prior
distribution:

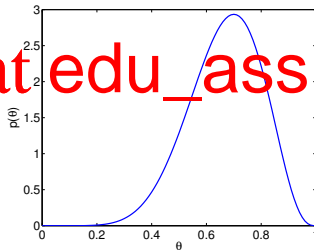
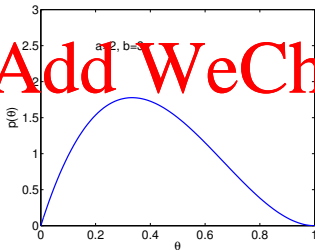
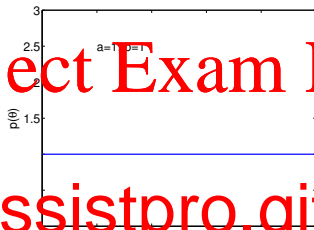
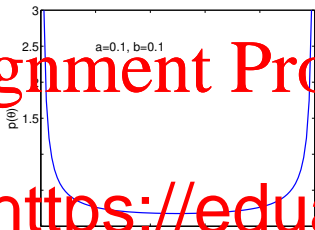
$$\text{Beta}(\theta|a, b) = \frac{1}{Z(a, b)} \theta^{a-1} (1-\theta)^{b-1}$$

where $Z(a, b)$ is a suitable normaliser

We can tune a, b to reflect our belief in the range of likely values of θ

Beta Prior

Examples



Beta Prior and Binomial Likelihood:

Beta Posterior Distribution

Recall that for $\mathcal{D} = \{x_1, \dots, x_N\}$, the likelihood under a Bernoulli model is:

Assignment Project Exam Help

where m

def

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

Beta Prior and Binomial Likelihood:

Beta Posterior Distribution

Recall that for $\mathcal{D} = \{x_1, \dots, x_N\}$, the likelihood under a Bernoulli model is:

Assignment Project Exam Help

where m

For the pri

<https://eduassistpro.github.io>

Add WeChat edu_assist_pro

$$\begin{aligned} p(\theta|\mathcal{D}, a, b) &= \frac{p(\mathcal{D}|\theta)}{\int_0^1 p(\mathcal{D}|\theta) d\theta} \\ &= \text{Beta}(\theta|m+a, \ell+b). \end{aligned}$$

Beta Prior and Binomial Likelihood:

Beta Posterior Distribution

Recall that for $\mathcal{D} = \{x_1, \dots, x_N\}$, the likelihood under a Bernoulli model is:

Assignment Project Exam Help

where m

For the pri

<https://eduassistpro.github.io>

Add WeChat edu_assist_pr

$$\begin{aligned}
 p(\theta|\mathcal{D}, a, b) &= \frac{p(\mathcal{D}|\theta)}{\int_0^1 p(\mathcal{D}|\theta) d\theta} \\
 &= \text{Beta}(\theta|m+a, \ell+b).
 \end{aligned}$$

Can use this as our new prior if we see more data!

Beta Prior and Binomial Likelihood:

Beta Posterior Distribution

Now suppose we choose θ_{MAP} to maximise $p(\theta|\mathcal{D})$

(MAP= Maximum A Posteriori)

One can show that

<https://eduassistpro.github.io>

cf. the esti

Add WeChat $\theta_{\text{ML}} = \frac{m}{N}$ edu_assist_pr

The prior parameters a and b can be seen a

What values of a and b ensure $\theta_{\text{MAP}} = \theta_{\text{ML}}$? $a = b = 1$. Make sense?

(Note that the choice of the beta distribution was not accidental here — it is the “conjugate prior” for the binomial distribution.)

1 The Bernoulli Distribution

Assignment Project Exam Help

2 The Binomial Distribution

3 Para

<https://eduassistpro.github.io>

4 Bayesian Parameter Estimation

Add WeChat edu_assist_pr

5 Wrapping up

Assignment Project Exam Help

- Distributions involving binary random variables
 - ▶ Bernoulli distribution

<https://eduassistpro.github.io>

- ▶
- Bayesian inference: Full posterior on the par
 - ▶ Beta prior and binomial likelihood
- Reading: Mackay §23.1 and §23.5; Bis

Add WeChat edu_assist_pro

Next time

Assignment Project Exam Help

- Ent <https://eduassistpro.github.io>

Add WeChat edu_assist_pr