STAT 514 Lecture Notes

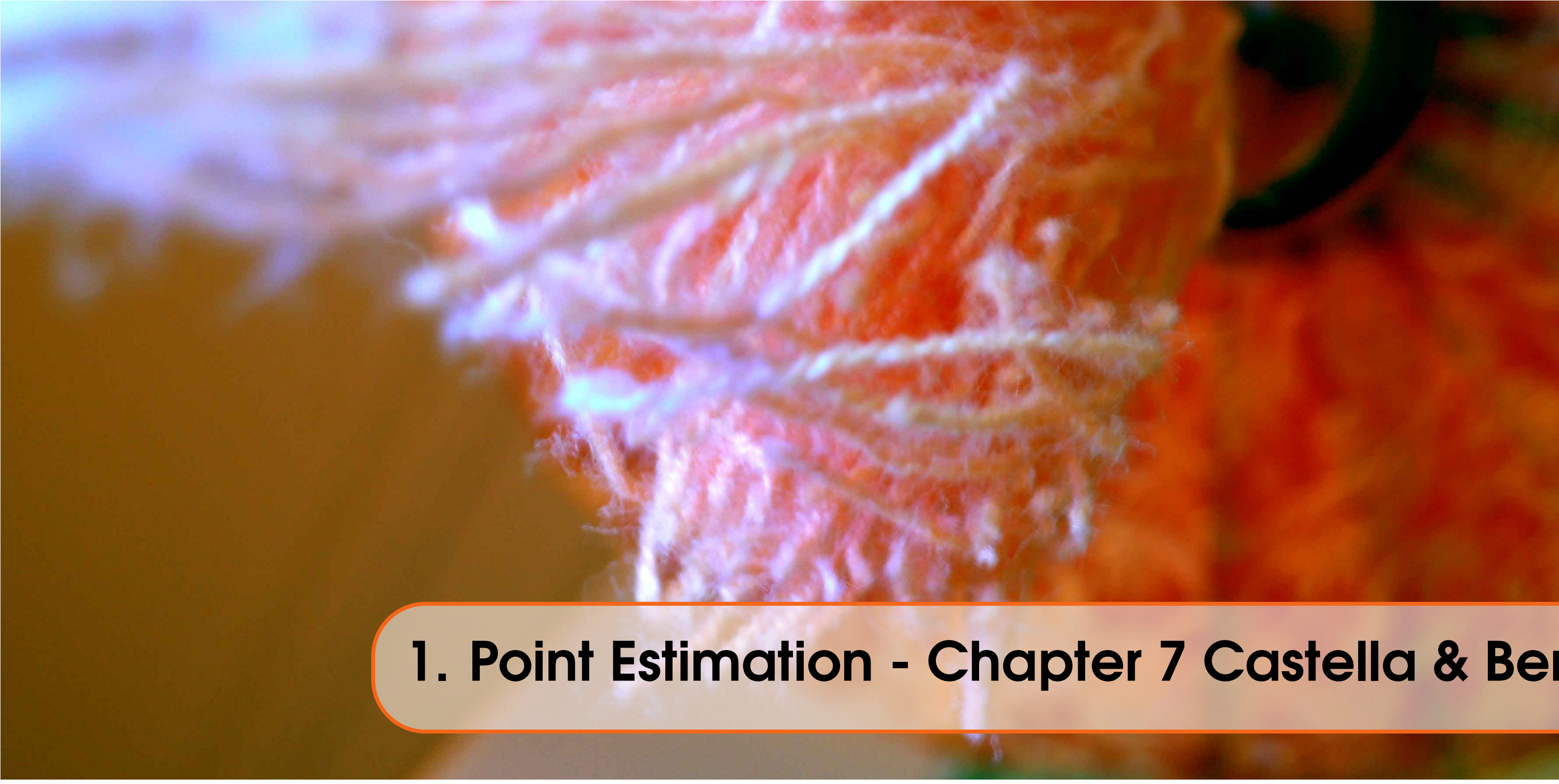
Dr. David Hunter

PUBLISHED BY PUBLISHER

BOOK-WEBSITE.COM

First printing, March 2013

Contents



1. Point Estimation - Chapter 7 Castella & Be

1.1 Introduction

In the simplest case, we have n observations of data that we believe follow the same distribution.

$$X_1, \ldots, X_n \overset{iid}{\sim} f_\theta(x)$$

where $f_\theta(x)$ is a density function involving a parameter θ. Our goal is to learn something about θ, which could be real or vector valued.

> **Definition 1.1.1 — Estimator.** An *estimator* of θ is any function $W(X_1, \ldots, X_n)$ of the data. That is, an estimator is a *statistic*.

Note:

1. $W(\boldsymbol{X})$ may not depend on θ.
2. $W(\boldsymbol{X})$ should resemble or "be close" to θ.
3. An estimator is *random*.
4. $W(X_1, \ldots, X_n)$ is the estimator, $W(x_1, \ldots, x_n)$ is the fixed estimate.

■ **Example 1.1** Suppose we have n observations from an exponential distribution,

$$X_1, \ldots, X_n \overset{iid}{\sim} f_\theta(x) = \frac{1}{\theta} \exp\left\{-\frac{x}{\theta}\right\} \mathbb{1}\{x > 0\}$$

for some $\theta > 0$. The **likelihood function** is equivalent to the joint density function, expressed as a function of θ rather than the data:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left\{-\frac{x}{\theta}\right\} = \frac{1}{\theta^n} \exp\left\{-\frac{1}{\theta} \sum_{i=1}^{n} x_i\right\}$$

This function represents the *likelihood* of observing the data we observed assuming the parameter was a particular value of θ. If we can maximize this function, we can determine the $\hat{\theta}$ for which the likelihood of observing $\boldsymbol{X}$ was the highest. This might tell us something about the true value of θ.

To maximize $L(\theta)$, we want to take the derivative, set it equal to 0, and solve for θ. However, in many cases taking the derivative of the likelihood function will be very hard, if not impossible.

We can use the fact that taking the logarithm does not change the location of extrema. The **log-likelihood function** in this case is

$$\ell(\theta) = \log L(\theta) = -n\log\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

Take the derivative with respect to the parameter and set equal to 0:

$$\begin{aligned} \ell'(\theta) &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i \overset{\text{set}}{=} 0 \\ \hat{\theta} &= \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned}$$

Here $\hat{\theta}$ is an estimator (the sample mean). Since it maximizes $L(\theta)$, we call it the **maximum likelihood estimator** (MLE). ■

1.2 Mean Squared Error

R Read Castella & Berger Chapter 7.3 - Methods of Evaluating Estimation

Definition 1.2.1 — Mean Squared Error. If $W(\boldsymbol{X})$ is an estimator of θ, then the **mean squared error** (MSE) is defined as

$$E_\theta\left[(W(\boldsymbol{X}) - \theta)^2\right].$$

Definition 1.2.2 — Unbiased estimator. If $W(\boldsymbol{X})$ is an estimator of θ, we say that $W(\boldsymbol{X})$ is **unbiased** if

$$E_\theta[W(\boldsymbol{X})] = \theta \quad \forall\theta.$$

Furthermore, the **bias** of $W(\boldsymbol{X})$ is

$$E_\theta[W(\boldsymbol{X})] - \theta.$$

■ **Example 1.2** For $\theta > 0$, let

$$X_1, \ldots, X_n \overset{iid}{\sim} f_\theta(x) = \theta x^{-2}\mathbb{1}\{x > \theta\}$$

Find the MLE of θ.

$$L(\theta) = \theta^n \prod_{i=1}^{n} x_i^{-2} \prod_{i=1}^{n} \mathbb{1}\{x > \theta\}$$

$$\hat{\theta} = \text{minimum of } x_i$$

■

Theorem 1.2.1 $\text{MSE}(W) = \text{bias}^2 + \text{Var}(W)$

Proof:

$$\begin{aligned} E[(W(\boldsymbol{X}) - \theta)^2] &= E[(W - E[W] + E[W] - \theta)^2] \\ &= E[(W - E[W])^2] + E[(E[W] - \theta)^2] + 2E[(W - E[W])(E[W] - \theta)] \\ &= \text{Var}(W) + \text{bias}^2(W) + 0 \end{aligned}$$

1.3 Best Unbias Estimator

What does best mean? Answer: Minimum variance.

Recall: Given an esitmator $W(\underline{X})$ for θ,

$$MSE(\theta) = E(W(\underline{X}) - \theta)^2)$$

Definition 1.3.1 — Best Unbiased Estimator. An estimator W* is a **best unbiased estimator*** of $\tau(\theta)$ if it satisties

$$E_\theta(W*) = \tau(\theta)$$

for all θ and, for any other estimator W with

$$E_\theta(W) = \tau(\theta)$$

we have

$$\text{Var}_\theta(W*) \leq \text{Var}_\theta(W)$$

for all θ. W* is also called a *uniform minimum variance unbiased estimator* (UMVUE) of $\tau(\theta)$.

1.4 Lost Function Optimality

Definition 1.4.1 — Loss. L(θ, W($\underline{x}$)) assigns a nonnegative real value called the **loss** to our decision to estimate θ by W($\underline{X}$). General context: Decision Theory.

Typically L(θ,θ) = 0 because nothing is lost if your decision is exactly correct.

■ **Example 1.3**

$$\begin{aligned} L(\theta, W(\underline{X})) &= (\theta - W(X)^2) \quad \text{square error loss} \\ &= |\theta - W(X)| \quad \text{absolute error loss} \\ &= \frac{W(X)}{\theta} - 1 - \log\frac{W(X)}{\theta} \quad \text{Stein's loss} \end{aligned}$$

■

Definition 1.4.2 — Risk. **Risk** of estimating θ by W($\underline{X}$) is

$$R(\theta, W) = E(\theta, W(\underline{X}))$$

Exercise 1.1 If $X_1, \ldots, X_n$ are iid with mean μ and varience σ^2 what is

$$E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)?$$

Note:

1. $\sum i = i^n \left(X_i - \bar{X})^2\right) = \sum(X_i^2) - n\bar{X}$

2. Var(X) = $E(X^2) - E(X)^2$
3. Var$(\frac{X_i}{n}) = \frac{\sigma^2}{n^2}$
4. Var$(\bar{X}) = \frac{\sigma^2}{n}$

Thus,

$$E(\sum X_i^2) = (\sigma^2 + \mu^2)n$$

$$-nE(\bar{X}^2) = -n(\mu^2 + \frac{\sigma^2}{n})$$

We may conclude, $\frac{1}{n-1}\sum(X_i - \bar{X})^2$ is an unbiased estimator of σ^2 called S^2. ■

Theorem 1.4.1 — Cauchy-Scwarz Inequality.

$$|| < x,y > ||^2 \leq || < x,x > || * || < y,y > ||$$

In terms of E(X) if A & B have $\mu = 0$:

$$(E(AB))^2 \leq E(A^2) * E(B^2)$$

Or in terms of covariance:

$$Cov^2(AB) \leq Var(A) * Var(B)$$

1.5 Citation

This statement requires citation [Smi12]; this one is more specific [Smi13, page 122].

1.6 Lists

Lists are useful to present information in a concise and/or ordered way[1].

1.6.1 Numbered List

1. The first item
2. The second item
3. The third item

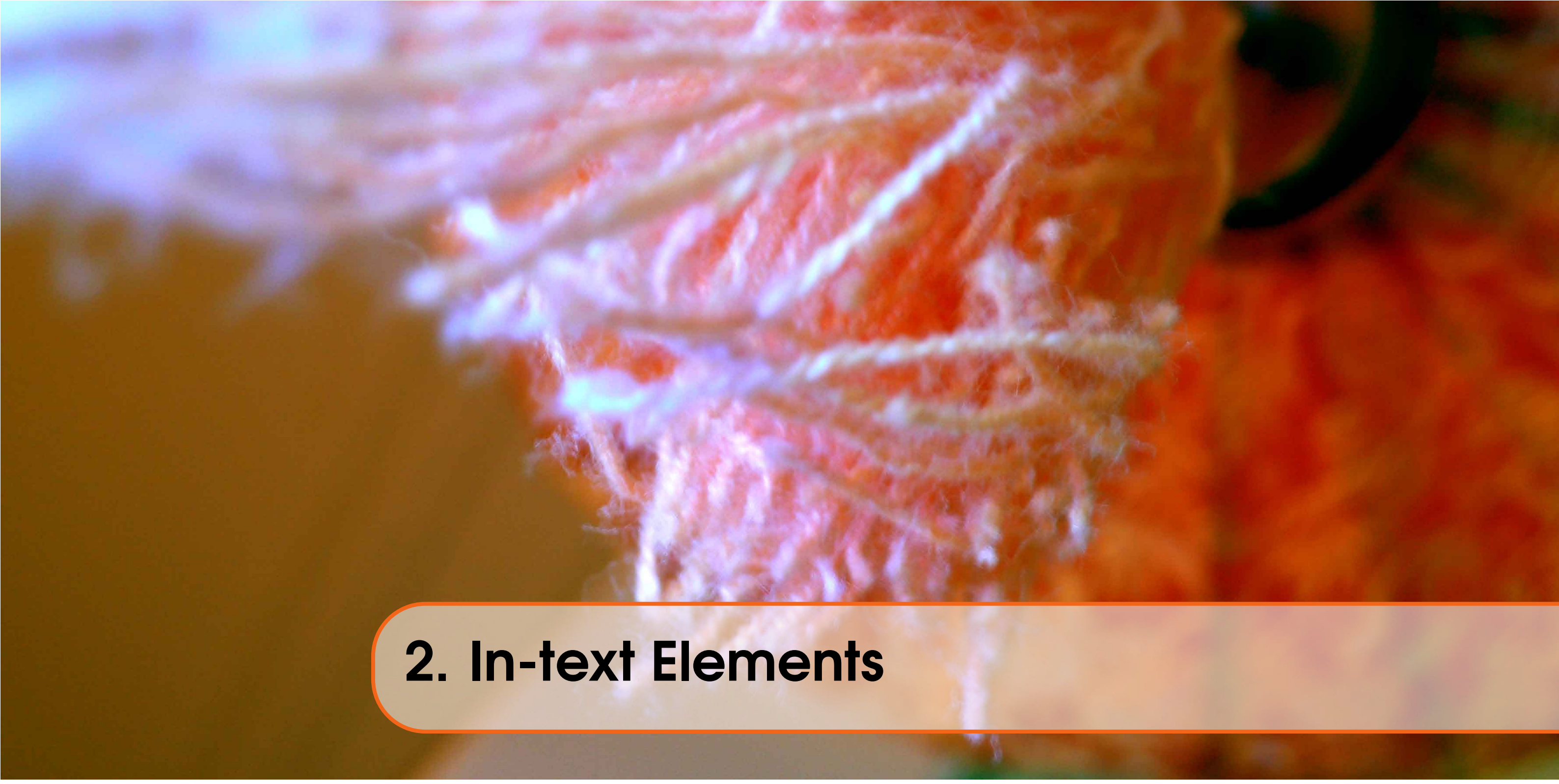
1.6.2 Bullet Points

- The first item
- The second item
- The third item

1.6.3 Descriptions and Definitions

Name Description
Word Definition
Comment Elaboration

[1] Footnote example...



2. In-text Elements

2.1 Theorems

This is an example of theorems.

2.1.1 Several equations

This is a theorem consisting of several equations.

Theorem 2.1.1 — Name of the theorem. In $E = \mathbb{R}^n$ all norms are equivalent. It has the properties:

$$\big| ||\mathbf{x}|| - ||\mathbf{y}|| \big| \leq ||\mathbf{x} - \mathbf{y}|| \tag{2.1}$$

$$||\sum_{i=1}^{n} \mathbf{x}_i|| \leq \sum_{i=1}^{n} ||\mathbf{x}_i|| \quad \text{where } n \text{ is a finite integer} \tag{2.2}$$

2.1.2 Single Line

This is a theorem consisting of just one line.

Theorem 2.1.2 A set $\mathscr{D}(G)$ in dense in $L^2(G)$, $|\cdot|_0$.

2.2 Definitions

This is an example of a definition. A definition could be mathematical or it could define a concept.

Definition 2.2.1 — Definition name. Given a vector space E, a norm on E is an application, denoted $||\cdot||$, E in $\mathbb{R}^+ = [0, +\infty[$ such that:

$$||\mathbf{x}|| = 0 \Rightarrow \mathbf{x} = \mathbf{0} \tag{2.3}$$

$$||\lambda \mathbf{x}|| = |\lambda| \cdot ||\mathbf{x}|| \tag{2.4}$$

$$||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \tag{2.5}$$

2.3 Notations

Notation 2.1. *Given an open subset G of $\mathbb{R}^n$, the set of functions φ are:*

1. *Bounded support G;*
2. *Infinitely differentiable;*

a vector space is denoted by $\mathscr{D}(G)$.

2.4 Remarks

This is an example of a remark.

R The concepts presented here are now in conventional employment in mathematics. Vector spaces are taken over the field $\mathbb{K} = \mathbb{R}$, however, established properties are easily extended to $\mathbb{K} = \mathbb{C}$.

2.5 Corollaries

This is an example of a corollary.

Corollary 2.5.1 — Corollary name. The concepts presented here are now in conventional employment in mathematics. Vector spaces are taken over the field $\mathbb{K} = \mathbb{R}$, however, established properties are easily extended to $\mathbb{K} = \mathbb{C}$.

2.6 Propositions

This is an example of propositions.

2.6.1 Several equations

Proposition 2.6.1 — Proposition name. It has the properties:

$$\big| ||\mathbf{x}|| - ||\mathbf{y}|| \big| \le ||\mathbf{x} - \mathbf{y}|| \tag{2.6}$$

$$|| \sum_{i=1}^{n} \mathbf{x}_i || \le \sum_{i=1}^{n} ||\mathbf{x}_i|| \quad \text{where } n \text{ is a finite integer} \tag{2.7}$$

2.6.2 Single Line

Proposition 2.6.2 Let $f, g \in L^2(G)$; if $\forall \varphi \in \mathscr{D}(G)$, $(f, \varphi)_0 = (g, \varphi)_0$ then $f = g$.

2.7 Examples

This is an example of examples.

2.7.1 Equation and Text

■ **Example 2.1** Let $G = \{x \in \mathbb{R}^2 : |x| < 3\}$ and denoted by: $x^0 = (1, 1)$; consider the function:

$$f(x) = \begin{cases} e^{|x|} & \text{si } |x - x^0| \le 1/2 \\ 0 & \text{si } |x - x^0| > 1/2 \end{cases} \tag{2.8}$$

The function f has bounded support, we can take $A = \{x \in \mathbb{R}^2 : |x - x^0| \le 1/2 + \varepsilon\}$ for all $\varepsilon \in]0; 5/2 - \sqrt{2}[$. ■

2.7.2 Paragraph of Text

■ **Example 2.2 — Example name.** Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris. ■

2.8 Exercises

This is an example of an exercise.

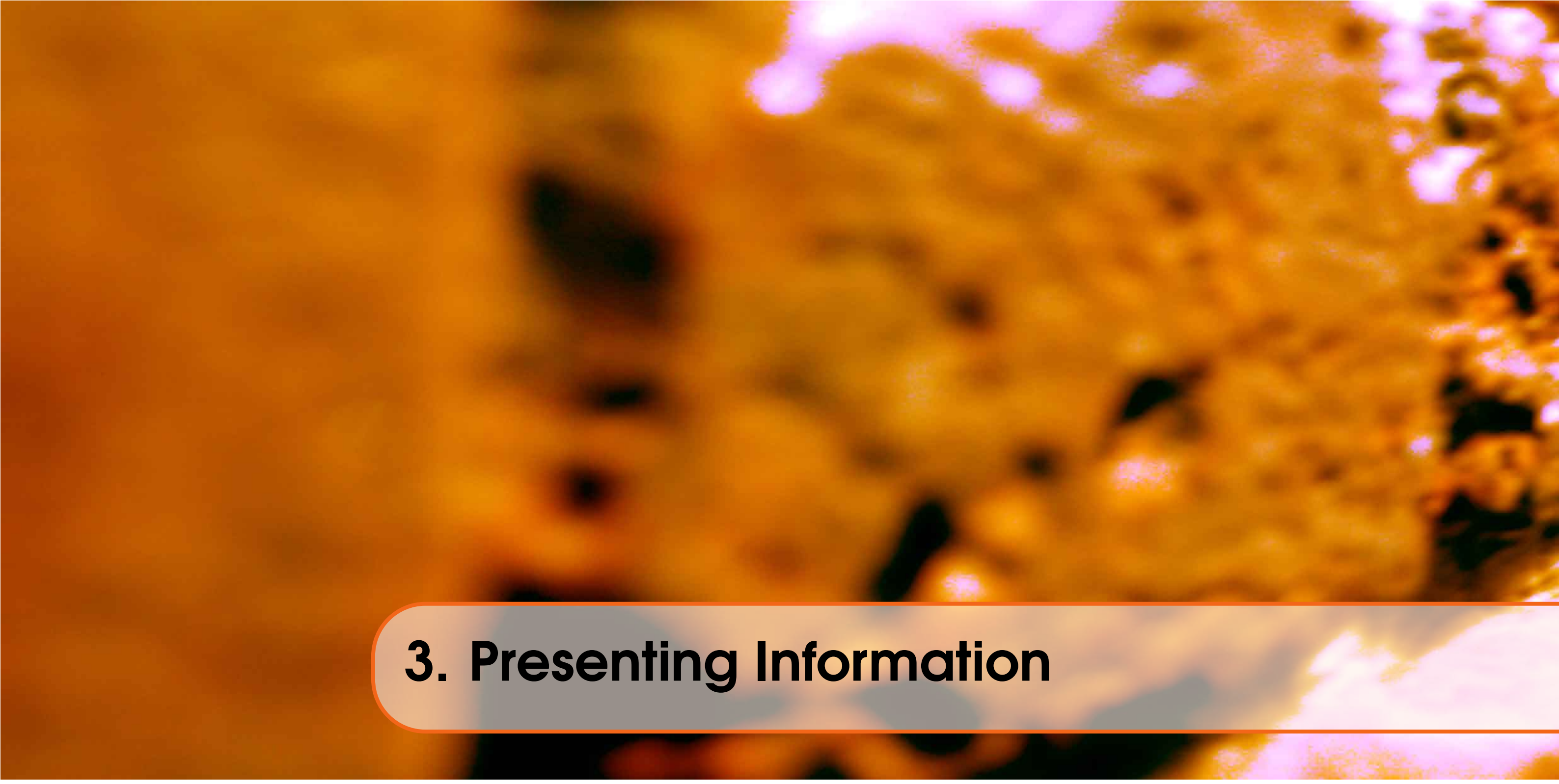
Exercise 2.1 This is a good place to ask a question to test learning progress or further cement ideas into students' minds. ■

2.9 Problems

Problem 2.1 What is the average airspeed velocity of an unladen swallow?

2.10 Vocabulary

Define a word to improve a students' vocabulary.
Vocabulary 2.1 — Word. Definition of word.



3. Presenting Information

3.1 Table

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table 3.1: Table caption

3.2 Figure

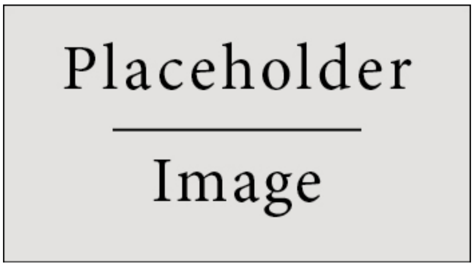


Figure 3.1: Figure caption

Bibliography

Books

[Smi12] John Smith. *Book title*. 1st edition. Volume 3. 2. City: Publisher, Jan. 2012, pages 123–200 (cited on page 8).

Articles

[Smi13] James Smith. “Article title”. In: 14.6 (Mar. 2013), pages 1–8 (cited on page 8).

Index