Maths Appendix
(for Agrifood Research)

UTAS, Draft 1.1,

January 17, 2022

Contents

1 A Little Bit of Mathematics

1.1 Basics

- **Law of Large Numbers**: As the sample size gets larger, sample statistics get ever closer to the population characteristics.
- **Central Limit Theorem**: Sample statistics computed from means (such as the means, themselves) are approximately normally distributed, (almost) regardless of the parent distribution. This approximation gets closer as the sample size increases.

1.2 Laws of Large Numbers

Large number theorems are about sample averages $\overline{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$.

Theorem 1 (Weak Law of Large Numbers). *If x_t are IID (identically and independently distributed) with finite mean μ then*

$$\overline{x}_T = \frac{x_1 + x_2 +, ..., +x_T}{T} \xrightarrow{p} \mu$$

Chebychev provides a modified Weak Law of Large Numbers in which the difference between a sample average and the true mean tends to zero, even for non IID data.

1.3 Central Limit Theorem

Theorem 2 (Central Limit Theorem). *If each x_t is a random draw from the same probability distribution with finite mean μ and finite variance σ^2 and $\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$ then*

$$\sqrt{T}(\bar{x}_T - \mu) \xrightarrow{d} N(0, \sigma^2).$$

The result is remarkable because it holds (almost) regardless of the form of the *pdf* for x_t. The distribution does not even need to remotely resemble a normal distribution.

1.4 Convergence

Definition 1 (Convergence in Probability). *The random variable x_t is said to "converge in probability" to c as $t \to \infty$, denoted by $x_t \xrightarrow{p} c$ or $\text{plim}(x_t) = c$, if*

$$\lim_{t\to\infty} \mathbb{P}(|x_t - c| > \varepsilon) = 0, \text{ for any } \varepsilon > 0.$$

Definition 2 (Convergence in Distribution). *The random variable x_t with distribution functions $F_t(x)$ is said to "converge in distribution" to a random variable x with distribution function $F(x)$ as $t \to \infty$, denoted by $x_t \xrightarrow{d} x$, if*

$$\lim_{t\to\infty} F_t(x) = F(x).$$

at every continuity point x of $F(x)$.

1.5 Delta Method

Loosely speaking, the delta method is an application of a Taylor series expansion for whatever function is used to estimate a parameter in question. Without loss of generality, assume a univariate case with a parameter estimate $\hat{\boldsymbol{\theta}}$ that is derived from some function g of a random statistic $\hat{z}$, that is

$$\hat{\theta} = g(\hat{z}), \tag{1}$$

where $\hat{z}$ has true mean μ_z and true variance σ_z^2. A Taylor series expansion of g around μ_z is

$$g(\hat{z}) = g(\mu_z) + g'(\mu_z)(\hat{z} - \mu_z) + \text{remainder}, \tag{2}$$

where the apostrophe denotes the first derivative. From equation (2) we have the classic results that $\mathbb{E}[\hat{\theta}] \approx g(\mu_z)$ and

$$\text{var}(\hat{\theta}) \approx (g'(\mu_z))^2 \sigma_z^2. \tag{3}$$

The "delta method" is usually taken to mean plugging in sample estimates of μ_z and σ_z into (3) in order to estimate the variance of the parameter estimate $\hat{\theta}$.

2 Probability Distributions

Section Goals ...

- Define key probability distributions.
- Highlight the simple relationships between these distributions.

2.1 Analytic Distributions

Think of "analytic" probability distributions as those you can write down on a piece of paper using straight-forward mathematical notation. They are essentially models, in which there are parameters denoted by Greek letters and "data" which is in the form of random variables. We will assume that you know the appropriate definitions and required conditions of the following.

- Probability density functions "*pdf*".
- Cumulative density functions "*cdf*".

In applied statistics, testing model parameters and overall models requires you to be familiar with the details of at least the Normal distribution, the t-distribution, the F-distribution and the χ^2-distribution.

2.2 Definitions I

These definitions follow Wasserman (2010).

- The **sample space** Ω is the set of possible outcomes of an "experiment". Points ω within Ω are **outcomes**, realisations or elements. Subsets of Ω are called **events**. For example if you toss a coin twice then $\Omega = \{HH, HT, TH, TT\}$. The event that the first toss is heads is $A = \{HH, HT\}$.
- We assign a real number $\mathbb{P}(A)$ to every event A. This is called the **probability** of A.

The problem is then to link samples spaces with actual data — this is done with the concept of a random variable.

- A **random variable** is a mapping $X : \Omega \to \mathbb{R}$ that assigns a real number $X(\omega)$ to each outcome ω.
- For example, flip a coin ten times. Let $X(\omega)$ be the number of heads in the sample outcome ω. Assume that $\omega = HHTHHTHHTT$, then $X(\omega) = 6$.

2.3 Definitions II

Given a random variable X, we define the cumulative distribution function or *cdf* as follows.

- A **cumulative density function** is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = \mathbb{P}(X \leq x) \tag{4}$$

The *cdf*, sometimes written as simply F, effectively contains all the information about the random variable. We can now define a probability density function or *pdf*.

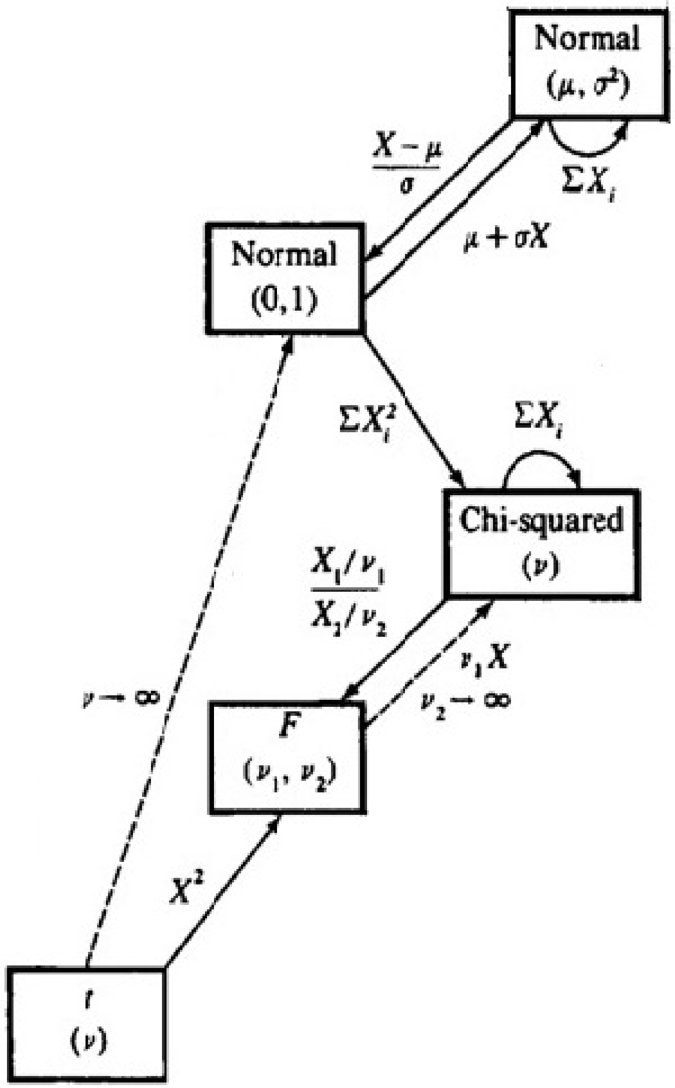
- A random variable X is **continuous** if there exists a function f_X such that $f_X(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} f_X(x)dx = 1$ and for every a and b,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx. \tag{5}$$

  The function f_X (or simply f) is called the **probability density function** or *pdf*.

We also have that $F_X(x) = \int_{\infty}^{x} f_X(t)dt$ and $f_X(x) = F'_X$ at all points at which F_X is differentiable.

2.4 PDF Map Snippet



The following relationships are useful for statistical testing.

1. If $X \sim N(\mu, \sigma^2)$ then $\frac{X-\mu}{\sigma} \sim N(0, 1)$.
2. The $X \sim N(0, 1)$ then $X^2 \sim \chi^2(1)$.
3. The sum of K variables that are $\chi^2(1)$ is $\chi^2(K)$.
4. If $X_1 \sim \chi^2(\nu_1)$ and $X_2 \sim \chi^2(\nu_2)$ then $\frac{X_1/\nu_1}{X_2/\nu_2} \sim F(\nu_1, \nu_2)$.
5. If $X \sim F(\nu_1, \nu_2)$ then $\nu_1 X \xrightarrow{d} \chi^2(\nu 1)$ as $\nu_2 \to \infty$.
6. If $X \sim t(\nu)$ then $X \xrightarrow{d} N(0, 1)$ as $\nu \to \infty$.
7. If $X \sim t(\nu)$ then $X^2 \sim F(1, \nu)$.

2.5 PDF Examples

2.5 PDF Examples

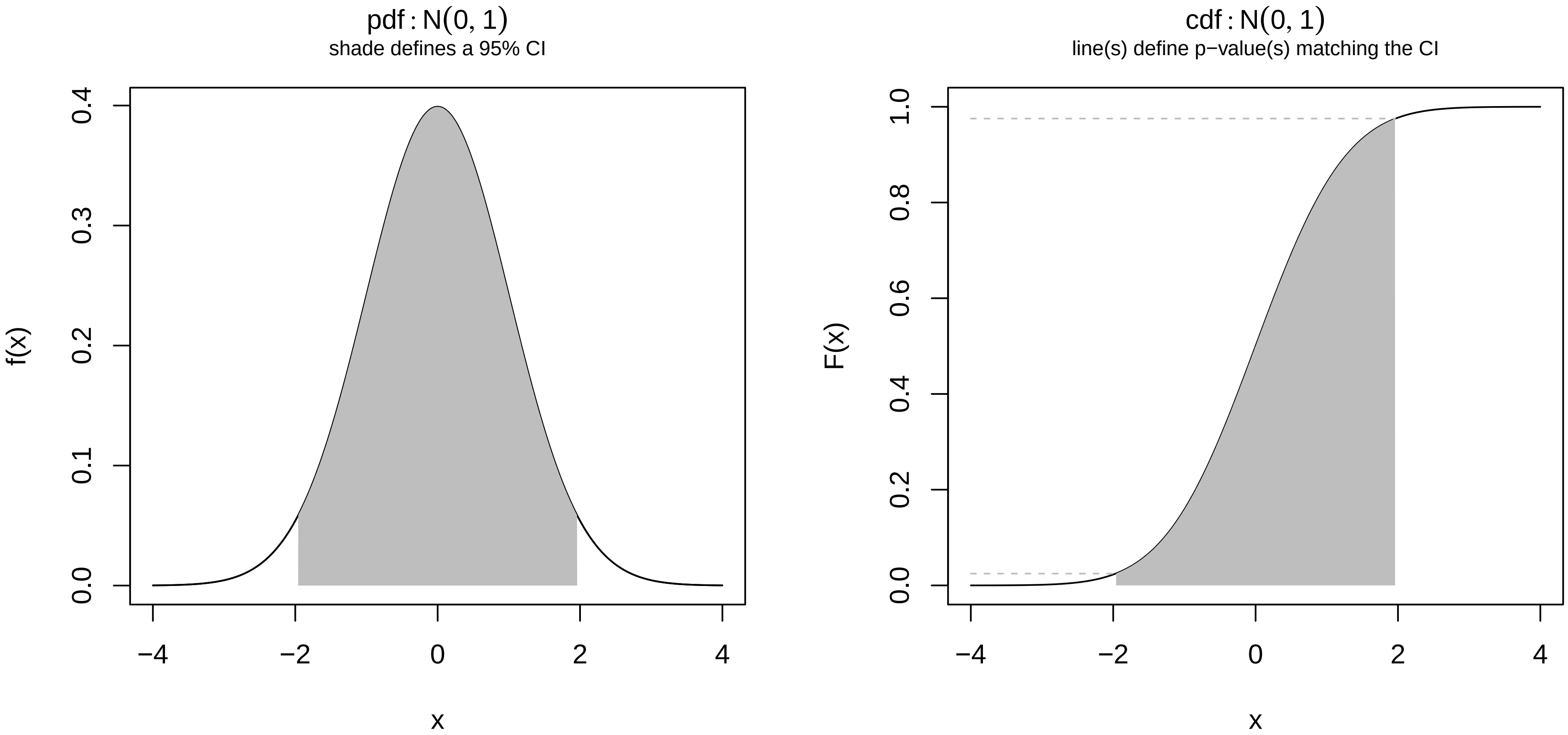


Figure 1: Example of a Normal *pdf* and *cdf*.

2.5 PDF Examples

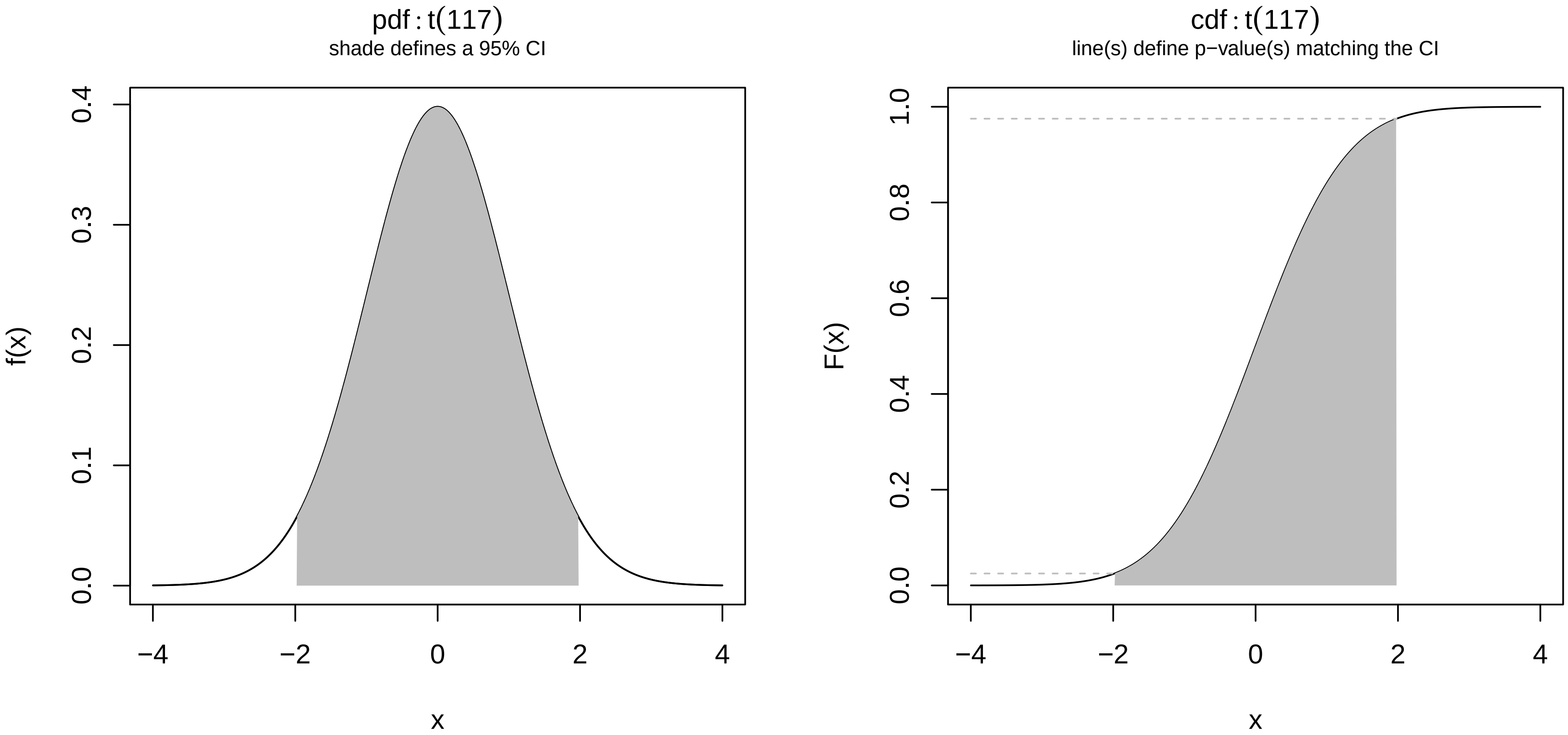


Figure 2: Example of a t-distribution *pdf* and *cdf*.

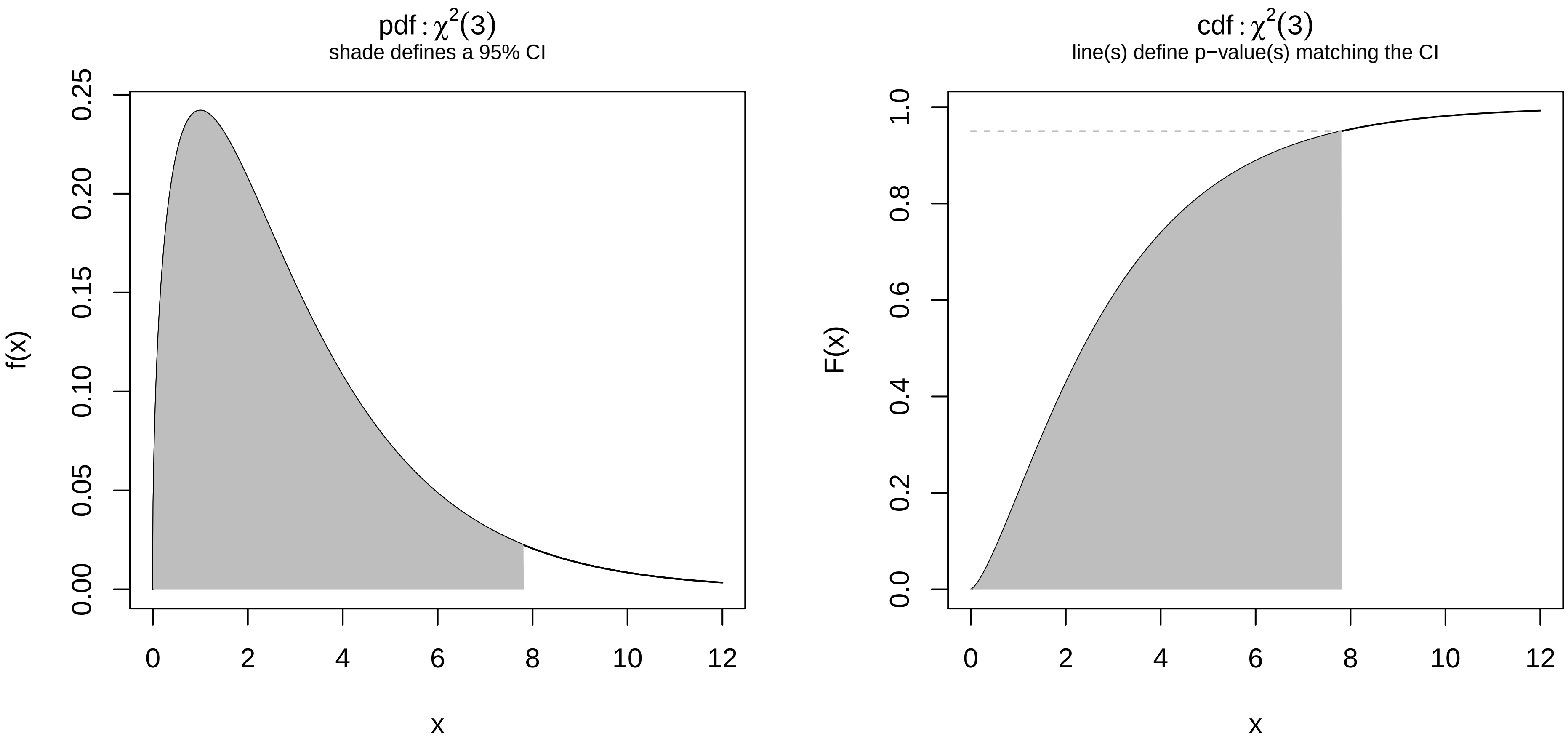


Figure 3: Example of a χ^2-distribution *pdf* and *cdf*.

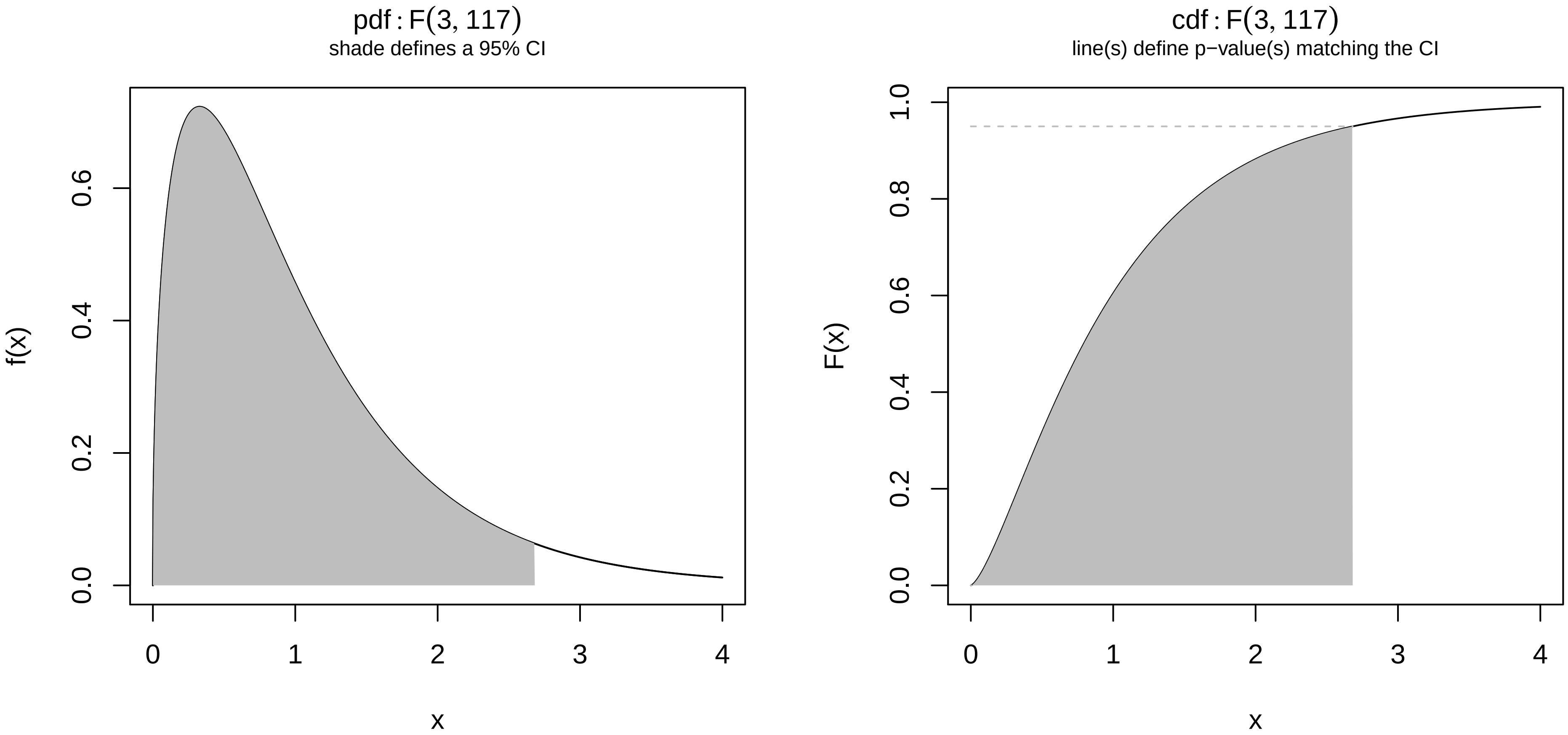


Figure 4: Example of an F-distribution *pdf* and *cdf*.

2.6 Empirical Distributions

Let an "empirical distribution" be a non-parametric "arrangement" of data that is a counterpart to an analytic distribution. Practical examples for univariate data sets include the following.

- An empirical *pdf* can be represented by a histogram (scaled to have an "area" of one under its bars).
- An empirical *cdf* is found by ordering the data, which can be conveniently plotted on the x-axis versus the interval [0,1] on the y-axis.

Advanced non-parametric techniques can characterize basic empirical distributions with more detail — though, most of the techniques are essentially approximating "join-the-dots" exercises.

Parametric model fitting can also add detail by replacing Greek letters with numbers. For example, you can assume the data in a histogram is from a Normal distribution and then use Maximum Likelihood Estimation to estimate μ and σ.

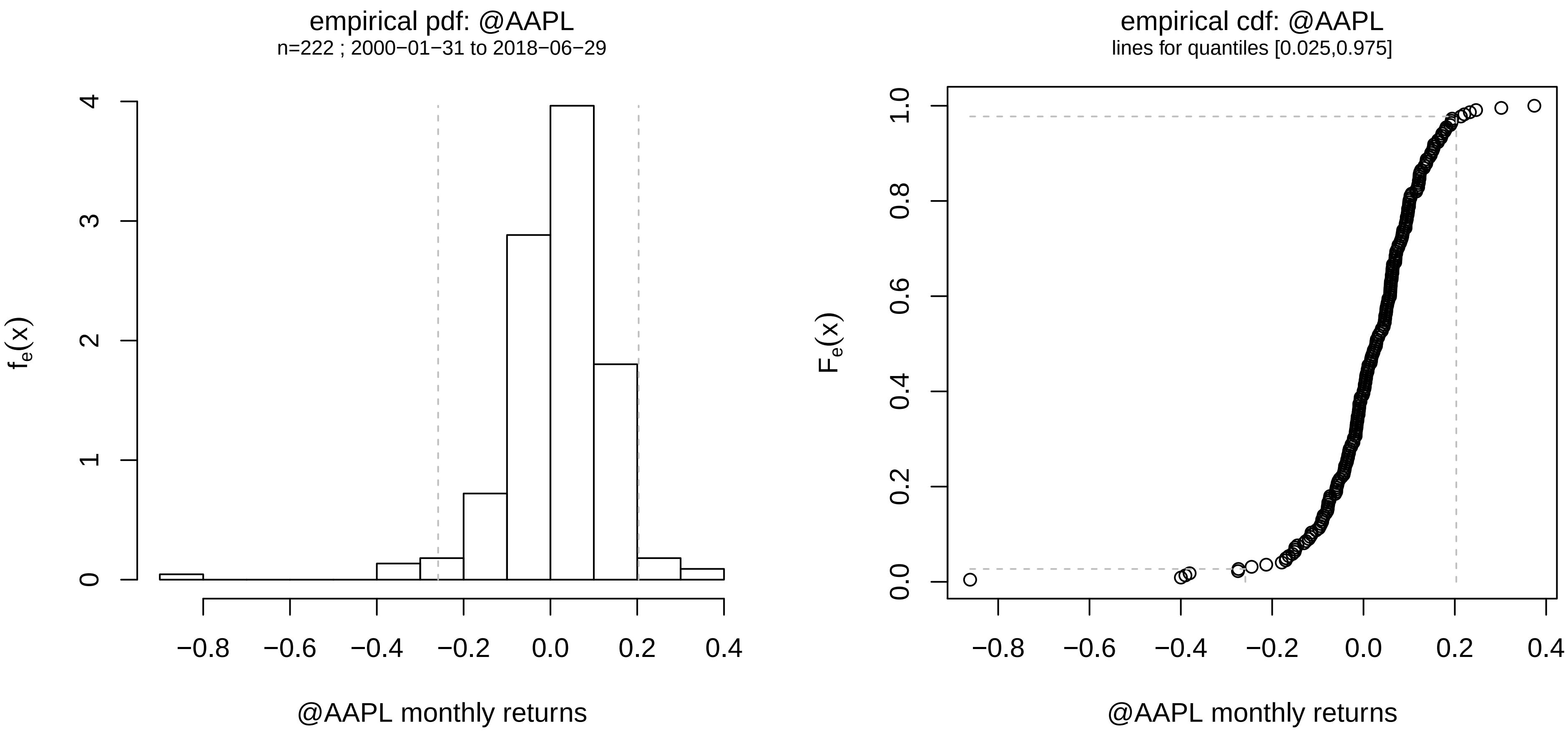


Figure 5: Empirical distribution for monthly Apple returns, since 2000.

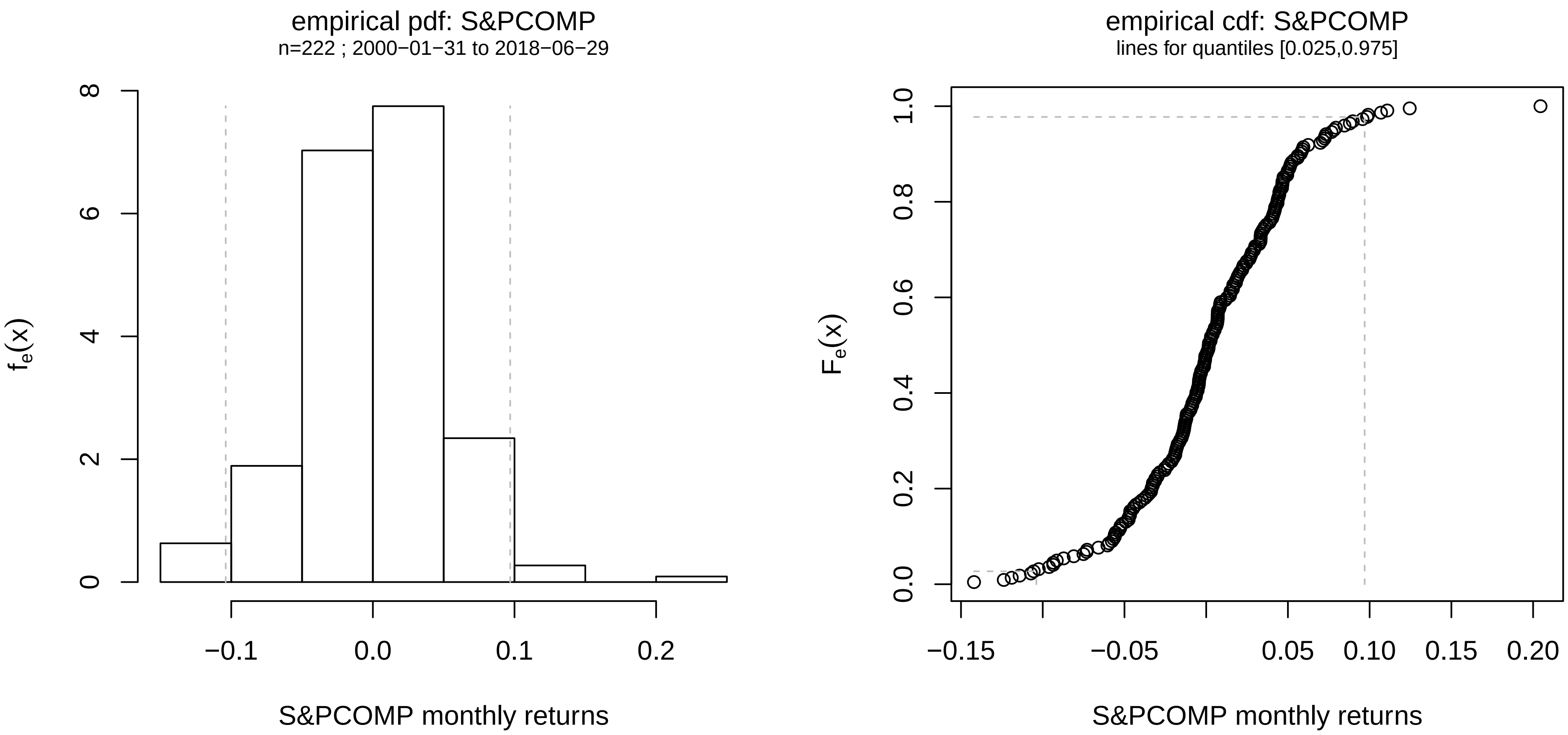


Figure 6: Empirical distribution for monthly S&P 500 index returns, since 2000.

3 References

References

Wasserman, L. (2010). *All of Statistics: A Concise Course in Statistical Inference*. Springer.