Unravelling the stellar Initial Mass Function of early-type galaxies with hierarchical Bayesian modelling

Dries, Matthijs

IMPORTANT NOTE: You are advised to consult the publisher’s version (publisher’s PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher’s PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1

Introduction
On a dark moonless night in the countryside one can discern a band of light that stretches across the sky. Already since the ancient Greeks, this band of light has been known as the Milky Way. It was only in 1610 that Galileo Galilei for the first time pointed a telescope at the Milky Way and discovered that it actually consisted of many faint stars that are not resolved by the naked eye.

More than a century later, the French comet hunter Charles Messier continuously ran into a set of fuzzy objects that were actually not comets because they did not move across the sky. To prevent him and his colleagues from wasting time on these objects in their search for comets, he decided to make a list of these objects. This list is now known as the Messier Catalogue. In the next century, many more of these objects were discovered, by, amongst others, the German-British astronomer William Herschel. This led to the publication of the extensive New General Catalogue (NGC) by John Dreyer in 1888.

The origin of the so-called “spiral nebulae” found in these catalogues has long been the subject of debate. Of particular interest in this discussion was the question whether these objects originated in the Milky Way or had an extragalactic origin. A noticeable highlight in the discussion was the 1920 Great Debate between Harlow Shapley and Heber Curtis on the “scale of the Universe”. It was, however, only in 1923 that the American astronomer Edwin Hubble finally resolved the debate. Using Cepheid variable stars, he managed to measure the distance to the Andromeda nebula and show that it is located far beyond the Milky Way. Nowadays, the Milky Way is known to be only one out of billions of other galaxies.

In the next decade Hubble continued his research into galaxies and developed a classification scheme for galaxies based on their visual appearance. This classification scheme is known as the Hubble sequence and divides galaxies into three main categories: ellipticals, lenticulars and spirals. Figure 1.1 gives a schematic overview of the Hubble sequence. Related to their location in the Hubble diagram, elliptical and lenticular galaxies are also referred to as early-type galaxies, whereas spirals are referred to as late-type galaxies.

Besides their morphological differences, early-type and late-type galaxies are also characterized by different colours. Late-type galaxies are in general bluish whereas early-type galaxies are yellowish-reddish. The different colours are related to the stars that constitute these galaxies. Spiral galaxies are actively forming stars and the most massive stars that are
formed are hot and blue. These massive stars outshine their less massive, cooler and redder counterparts, making the galaxy appear blue. Elliptical galaxies, on the other hand, have very little ongoing star formation. The massive stars that make spiral galaxies blue have relatively short lifetimes, and hence these stars are no longer present in ellipticals. Therefore, the light of ellipticals is dominated by the long-lived redder population of lower-mass stars, giving ellipticals their characteristic yellow-red colour.

Low-mass stars are thought to be much more abundant than high-mass stars. Within the Milky Way, star counts of individual stars have given us a reasonable idea of how the number of stars changes as a function of stellar mass (Salpeter 1955; Kroupa 2001; Chabrier 2003). For more distant galaxies, stars are unresolved and these kind of measurements are not possible. The distribution of stellar masses in these galaxies is therefore often assumed to be the same as in the Milky Way. But is the stellar mass distribution really the same in all galaxies? And how can one reliably measure the distribution of stellar masses in unresolved galaxies? These questions are the central theme of this thesis, where we develop a Bayesian model for inferring the stellar mass distribution of unresolved galaxies.

In the remaining of this chapter, I will give a general introduction into the formation and evolution of galaxies (Section 1.1), stellar population synthesis (Section 1.3), the initial mass function (IMF) (Section 1.4), and

---

**Figure 1.1** – Classification of galaxies in the Hubble tuning fork diagram. Image credit: Hubble (1936).
Bayesian model comparison (Section 1.5). The outline of this thesis will be given in Section 1.6.

1.1 Formation and evolution of galaxies

Nowadays, the standard model of cosmological structure formation is provided by the Λ cold dark matter (ΛCDM) paradigm (Liddle 2003). The ΛCDM paradigm is a cosmological model that seeks to explain the expansion of the Universe and the formation and evolution of the structures within it. The ΛCDM paradigm assumes that the Universe started with a single event: the Big Bang.

Within the ΛCDM model, the energy content of the Universe is made of three different components: radiation, matter and dark energy. The matter component can be divided into a baryonic component (i.e. the ordinary matter of which stars and planets are made) and a (cold) dark matter component. Dark matter only interacts through gravity and the weak interaction, and although there are several indications of its existence, the exact nature of it is currently unknown.

During the first three minutes after the Big Bang, the Universe was very hot and the hydrogen, helium and lithium atoms that formed during primordial nucleosynthesis remained completely ionized. This implies that the Universe is effectively opaque to electromagnetic radiation, since all photons were almost instantaneously scattered by free electrons and radiation and baryonic matter were in thermal equilibrium. As the Universe expanded it cooled down, and roughly 380,000 years after the Big Bang, the temperature became low enough for protons and electrons to form neutral hydrogen (the most abundant element in the Universe). This is referred to as recombination. After recombination, the Universe became basically transparent to radiation, and baryonic matter decoupled from radiation. A relic of this surface of last scattering is the Cosmic Microwave Background (CMB) discovered by Penzias & Wilson (1965).

The very early Universe after the Big Bang is thought to consist of an almost isotropic and homogeneous density field. This initial density field contained very tiny quantum fluctuations, which according to the inflation model (Guth 1981) were amplified enormously during a rapid phase of exponential expansion. These fluctuations are thought to be the seeds of the variety of structures that we nowadays observe in the Universe and are
confirmed by small fluctuations ($\sim 10^{-5}$) in the CMB temperature (Jarosik et al. 2011; Planck Collaboration et al. 2016).

The energy content of the Universe is nowadays dominated by dark energy (68.6%) and dark matter (26.8%), whereas the contribution of baryonic matter is much smaller (4.9%) and in terms of energy density photons are almost negligible (Planck Collaboration et al. 2016).

Before recombination, baryonic matter is coupled to radiation and radiation pressure prevents structure growth in the baryonic density field. However, since dark matter does not interact with photons, the small initial fluctuations in the dark matter density field start to grow under the influence of gravity. While the cosmic density $\rho_c$ is decreasing as a consequence of the expanding background, the density contrast $\delta = \rho/\rho_c - 1$ of overdense regions increases as it attracts matter from its surroundings.

Initially, structure growth is reasonably well described by linear perturbation theory (Peebles 1980). When the density contrast reaches a value of $\delta \approx 1$, it is no longer sufficient to describe the evolution of growing structures with linear perturbation theory and one has to use the full non-linear fluid equations. This phase is referred to as non-linear structure formation. Around the time that the density contrast reaches a value of $\delta \sim 1$, the overdensity reaches the point of turnaround, and it decouples from the expanding Universe and starts to collapse under the influence of its own gravity. The initial overdensity becomes a gravitationally bound object, which we refer to as a dark matter halo. When the dark matter halo has virialized, it has collapsed to approximately half its radius at turnaround.

After recombination, perturbations in the baryonic density field also start to grow under the influence of gravity. Since baryonic matter comprises only a relatively small fraction of the total mass in the Universe, it is expected to follow the dark matter distribution closely. The galaxies in our Universe are therefore thought to have formed in dark matter haloes.

According to the $\Lambda$CDM cosmology, structure formation is also hierarchical. In this scenario, the smallest-scale structures are the first to become non-linear and form dark matter haloes. As time progresses, these small dark matter haloes continuously merge to form more massive objects. The distribution of dark matter haloes in the Universe as a function of time may be derived from cosmological N-body simulations but there are also analytical formalisms such as the Press-Schechter formalism (Press & Schechter 1974).
As baryonic matter settles into hydrostatic equilibrium in dark matter haloes, it is able to lose energy through radiative cooling processes. At some point this may trigger Jeans instabilities (Jeans 1902) and lead to the formation of the first stars and galaxies. Since the infalling baryonic gas conserves its angular momentum, the first galaxies are expected to have a rotating disk shape.

Galaxies can grow through a series of subsequent merger events. Within that context, we distinguish between minor and major mergers. In the case of a minor merger, a small galaxy is “ingested” by a much larger galaxy, leaving the dynamical structure of the more massive progenitor largely intact. Massive disk galaxies such as the MW are thought to assemble their mass through a sequence of minor mergers. For a major merger, the galaxies that merge have comparable masses, and after a major merger the original dynamical structures of these galaxies are often destroyed and the newly formed object is dominated by random motions. Major mergers are thought to play an important role in the formation of elliptical galaxies. Another process that potentially plays an important role in the formation of galaxies are cold streams of gas that originate in the cosmic web and can provide fuel for star formation in a shock-heated dark matter halo (Birnboim & Dekel 2003; Dekel et al. 2009).

1.2 Elliptical galaxies

In this thesis we will focus on early-type galaxies (ETGs). The main reasons for this are that there are indications that ETGs have an IMF that is different from the MW IMF (see Section 1.4) and that the old stellar populations that characterize ETGs make it easier to infer their IMF.

Most of the galaxies in the Universe are spiral galaxies. Nevertheless, ETGs contain more than half of the stellar mass in the Universe (Bell et al. 2003; Gallazzi et al. 2008). ETGs are characterized by an ellipsoidal shape and a nearly smooth surface brightness distribution. For giant and midsized ellipticals the de Vaucouleurs’ law (de Vaucouleurs 1948) provides a good description of the surface brightness profile. The de Vaucouleurs’ profile is given by

\[
\ln I(R) = \ln I_0 - kR^{1/4},
\]

in which \(R\) is the radius from the center and \(I_0\) is the surface brightness at \(R = 0\). More generally, the surface brightness profile of an ETG can be
described by a Sérsic profile (Sérsic 1963),

$$\ln I(R) = \ln I_0 - kR^{1/n}, \quad (1.2)$$

which for $n = 4$ reduces to the de Vaucouleurs’ profile.

The main stellar component of an ETG is a population of old and red stars. Historically, ETGs are thought to contain little gas and dust and almost no ongoing star formation. However, there is evidence that some ETGs do contain gas and show signs of a small amount of residual star formation (Young et al. 2011; Crocker et al. 2011). In the densest environments of the Universe, elliptical galaxies are more common than in less dense environments. This is known as the morphology-density relation (Dressler 1980), and it provides important clues on the formation and evolution of galaxies.

There are several empirical scaling relations between the properties of elliptical galaxies. One of the most important scaling relations is the fundamental plane (Djorgovski & Davis 1987; Dressler et al. 1987), which relates the effective radius $R_{\text{eff}}$ (i.e. the radius at which half of the total light of the system is emitted), the central velocity dispersion $\sigma_c$ and the average surface brightness within the effective radius $\langle I_{\text{eff}} \rangle$ of the galaxy as

$$R_{\text{eff}} \propto \sigma_c^a \times \langle I_{\text{eff}} \rangle^b. \quad (1.3)$$

By measuring the velocity dispersion $\sigma_c$ and the average surface brightness within the effective radius $\langle I_{\text{eff}} \rangle$, the fundamental plane can be used to determine the effective radius $R_{\text{eff}}$. The physical size of the effective radius is not directly observable, but by combining it with the angular size of the galaxy, the fundamental plane allows one to determine distances to ETGs.

### 1.3 Stellar population synthesis

In the study of unresolved galaxies, spectra or spectral energy distributions of these galaxies are an essential source of information. Many of the physical processes that take place in these galaxies leave an imprint on the spectra that we observe. Modelling these spectra with a stellar population synthesis (SPS) model allows us to study the physical processes in these galaxies and to infer properties of these galaxies that are not directly observable. In this section I will give a basic overview of SPS models. For a more complete
overview I refer the reader to the excellent reviews by Walcher et al. (2011) and Conroy (2013).

### 1.3.1 Single Stellar Populations

A single stellar population (SSP) is a population of stars with the same age and the same chemical composition. Creating a model spectrum of an SSP required three basic ingredients: stellar evolution (isochrones), a stellar library and an IMF. An isochrone of a given age and metallicity provides the stellar parameters of the stars that are present in an SSP. When we combine these stellar parameters with a stellar library, this allows us to create a spectrum for each of the stars in the isochrone. Finally, the IMF prescribes the number of stars that we have for each of the stars in the isochrone. Multiplying each of the isochrone spectra with the number of stars that are present in an SSP according to the IMF allows us to create a model spectrum for the SSP. This process is schematically shown in Figure 1.2. Mathematically, the construction of an SSP model spectrum can be summarized as

$$f_{\text{SSP}} = \int_{m_{\text{low}}}^{m_{\text{up}(t)}} f_{\text{star}}[T_{\text{eff}}(M), \log g(M) \mid t, Z] \xi(M) dM,$$

(1.4)
in which \( f_{\text{star}}[T_{\text{eff}}(M), \log g(M)|t, Z] \) is the spectrum of a star with effective temperature \( T_{\text{eff}} \) and surface gravity \( \log g(M) \) for a given age \( t \) and metallicity \( Z \). The IMF in the above relation is represented by \( \xi(M) \equiv dN/dM \) and the integral is evaluated from the lowest mass stars \( m_{\text{low}} \) to the (time-dependent) highest mass stars \( m_{\text{up}}(t) \) (this upper limit is set by our knowledge of stellar evolution). Usually the lower limit is set to the hydrogen burning limit (\( m_{\text{low}} \approx 0.1 \, \text{M}_\odot \)) and the initial upper limit is taken to be \( m_{\text{up}} \approx 100 \, \text{M}_\odot \).

The relation between the surface gravity, mass and effective temperature for a stellar population of a certain age and metallicity is defined by isochrones. The individual stellar spectra \( f_{\text{star}}[T_{\text{eff}}(M), \log g(M)|t, Z] \) are taken from either theoretical models or empirical stellar libraries. We now discuss the isochrones and stellar libraries in more detail. The IMF is discussed in more detail in Section 1.4.

With numerical models of stellar evolution it is possible to calculate the evolutionary track of a star with a given mass and metallicity \( Z \). An evolutionary track describes how a star evolves in the Hertzsprung-Russel (HR) diagram as a function of age. By calculating evolutionary tracks for a wide range of initial masses, e.g. \( M = 0.1 - 100 \, \text{M}_\odot \), one can construct a library of evolutionary tracks. Evolving each of these tracks for a given metallicity \( Z \) to the same age \( t \) and combining the parameters of these tracks gives us an isochrone of age \( t \) and metallicity \( Z \). One such isochrone defines the stellar parameters of the stars that are present in an SSP.

Nowadays, different sets of homogeneous evolutionary tracks and isochrones are available. Examples are the Parsec isochrones (Bertelli et al. 1994; Girardi et al. 2000; Marigo et al. 2008), the Geneva isochrones (Schaller et al. 1992; Meynet & Maeder 2000), the Dartmouth isochrones (Dotter et al. 2008), the BaSTI isochrones (Pietrinferni et al. 2004; Cordier et al. 2007) and the Lyon isochrones (Chabrier & Baraffe 1997; Baraffe et al. 1998). These isochrones may have different assumptions with respect to the input physics that was used to calculate the evolutionary tracks, and there are several uncertainties that may eventually affect the SSP model spectra that are created with these isochrones. Among these uncertainties are the use of one-dimensional stellar evolution codes, the adopted level of convective overshooting, the modelling of stellar rotation, the interaction of close binary stars, mass loss, and the modelling of the thermally pulsating asymptotic giant branch (TPAGB). Since there is not a single isochrone model that covers all stellar masses, metallicities and evolutionary stages,
it is quite common to combine different sets of isochrones. However, this is not straightforward due to the different assumptions made in different models.

To subsequently convert the stellar parameters of an isochrone into a set of stellar spectra requires the use of a stellar library. Stellar libraries can be either theoretical or empirical. Theoretical libraries (e.g. Kurucz 1992; Coelho et al. 2005; Husser et al. 2013) have the advantage of providing a dense coverage of parameter space and having a high spectral resolution. Important uncertainties in theoretical spectra are the treatment of convection, incomplete lists of atomic and molecular lines, uncertain strengths and central wavelengths of molecular and atomic lines, and insufficient knowledge of the molecular partition function. Since empirical libraries are based on real stars, they by definition suffer less from these uncertainties. The main uncertainties of empirical libraries are determined by observational constraints. Moreover, empirical libraries have the disadvantage of not completely covering the parameter space of the isochrones (especially for the late evolutionary stages and for lower metallicities). Another source of uncertainty for empirical libraries is the determination of the stellar parameters of the stars in the library. Examples of empirical libraries are STELIB (Le Borgne et al. 2003), ELODIE (Prugniel & Soubiran 2001, 2004; Prugniel et al. 2007), INDO-US (Valdes et al. 2004), MILES (Sánchez-Blázquez et al. 2006), IRTF (Rayner et al. 2009) and XSL (Chen et al. 2011).

1.3.2 Composite Stellar Populations

Modelling the spectrum of a real galaxy is further complicated by the fact that a galaxy is not an SSP but has an extended star formation history (SFH). Therefore, a galaxy is a composite stellar population (CSP). Moreover, the spectrum of a galaxy that we observe is affected by dust.

There are various indicators that may be used to probe the star formation rate of a galaxy. Examples of these indicators are the total UV luminosity, infra-red indicators based on the processing of starlight by dust and emission lines from ionized gas (Kennicutt & Evans 2012). However, to create an accurate model for the spectrum of a CSP requires an integration over the SFH of the CSP. Classically, the SFH of a galaxy has been assumed to be exponentially declining, resulting in so-called $\tau$-models for which $\text{SFR} \propto e^{-t/\tau}$ (Papovich et al. 2001; Shapley et al. 2005). Other
parameterizations of the SFH include e.g. inverted $\tau$-models (Maraston et al. 2010; Pforr et al. 2012) and delayed $\tau$-models for which $\text{SFR} \propto t e^{-t/\tau}$ (Sandage 1986; Lee et al. 2010). Besides these simple parameterizations it is also possible to use a library of SFHs from hydrodynamical simulations or semi-analytical models of galaxy formation (Finlator et al. 2007; Pacifici et al. 2012) or to use nonparametric SFHs (Ocvirk et al. 2006).

The observed spectrum of a galaxy can be affected by dust in a number of ways. First of all, dust along the line of sight to the observed galaxy causes extinction. An observed spectrum may be corrected for this by applying an extinction correction (Cardelli et al. 1989; Calzetti 2001). Secondly, in galaxies that contain dust, light at shorter wavelengths may be absorbed by the dust and re-emitted at longer wavelengths (in particular the infra-red). For ETGs this is in general not considered to be an important process because the amount of dust in ETGs is relatively low. Finally, asymptotic giant branch (AGB) stars are characterized by mass-loss and this mass-loss can potentially be dust-rich (Bedijn 1987) and therefore affect the spectra of AGB stars. It is not straightforward to model this process and most SPS models do not take it into account. Nevertheless, it might be important because AGB stars can contribute significantly to the total luminosity of a galaxy (Kelson & Holden 2010; Villaume et al. 2015).

Taking into account the SFH, the model spectrum $f_{\text{CSP}}$ of a CSP may be written as

$$f_{\text{CSP}}(t) = \int_{t' = 0}^{t} \int_{Z = 0}^{Z_{\text{max}}} \text{SFR}(t - t') \text{P}(Z, t - t') f_{\text{SSP}}(t', Z) \, dt' \, dZ. \quad (1.5)$$

In this equation, $\text{SFR}(t - t')$ is the SFR at time $t - t'$ representing the SFH, $\text{P}(Z, t - t')$ is the time-dependent metallicity distribution function, and $f_{\text{SSP}}$ is the spectrum of an SSP as calculated in Equation 1.4. Note that this equation neglects the effects of dust attenuation on the observed galaxy spectrum.

1.3.3 Population Synthesis models

Among the first population synthesis models were the models developed by Tinsley (1968), Spinrad & Taylor (1971) and Faber (1972). The models of Spinrad & Taylor (1971) and Faber (1972) tried to determine the individual contributions of stars in the HR diagram to integrated spectra but suffered from a lack of astrophysical constraints. Tinsley (1968) introduced the idea
of evolutionary populations synthesis in which theories of stellar evolution
are used to put constraints on the available range of stellar templates in
the HR diagram.

Nowadays, a variety of different population synthesis models have been
developed (e.g. Bruzual & Charlot 2003; Le Borgne et al. 2003; Maraston
2005; Conroy & van Dokkum 2012a; Vazdekis et al. 2015) to infer the
properties of unresolved galaxies. Every model has its own set of ingredients
and parameters. Examples of such variety includes the isochrones that are
used, the stellar library that is used, the parameterization of the IMF
and modelling the spectrum as an SSP or not. The SFH of ETGs is
expected to be different than that of the Milky Way, and as a consequence
the chemical composition of ETGs is expected to be non-solar. Some
population synthesis models take this effect into account by using response
functions for variable abundance patterns. Another important aspect of a
population synthesis model is how the models are fitted to the data. In
what we described so far, the main focus has been the construction of a
model spectrum for an SSP or CSP. Yet, ultimately we would like to use
these models to extract information from observed galaxy spectra which
requires some sort of a fitting routine.

1.4 The initial mass function

The IMF is a probability distribution that describes the distribution of
stellar masses of the stars that form in a star formation event. The first
attempt to measure the IMF was made by Salpeter (1955). Although often
referred to as measuring the IMF, most of the time it is actually the present-
day mass function (PDMF) that is measured. The PDMF includes the
effects of stellar evolution and the SFH.

Salpeter (1955) derived the IMF for stars with masses between 0.4 and
10 \( M_\odot \) in the solar neighbourhood and determined that it could be described
with a single power law as

\[
\xi(M) \equiv \frac{dN}{dM} \propto M^{-\alpha},
\]

with \( \alpha = 2.35 \). Later on it was shown that for the MW below 0.5 \( M_\odot \)
the IMF becomes flatter and that a broken power law (Kroupa et al. 1993;
Kroupa 2001) or a combination of a power law with a lognormal distribution
(Chabrier 2003) provides a more accurate description of the MW IMF. In
1.4. The initial mass function

fact, these two distributions can be very similar and hard to distinguish, as shown by Dabringhausen et al. (2008).

Measurements of the IMF in the MW are mostly based on star counts. These measurements are complicated by the different volumes that are probed for different stellar masses, by the degeneracy with the SFH of the MW (Elmegreen & Scalo 2006) and by the effect of binaries on the inferred stellar mass spectrum (Metchev & Hillenbrand 2009; Goodwin & Kouwenhoven 2009). Studies of the IMF in different environments of the MW seem to indicate that the Galactic IMF is universal (Bastian et al. 2010).

For galaxies beyond the Local Group, direct star counts are not possible any more because the light of these galaxies is spatially unresolved. To determine the IMF of these systems one has to use different techniques. One of these techniques is the analysis of the spectrum of such a galaxy with a population synthesis model. Inferring the IMF and in particular the low-mass end of the IMF from the spectrum of a galaxy is not straightforward, however. Dwarfs with masses below 0.5 M_☉ contribute only a few percent to the integrated light of an old SSP. Nevertheless, these dwarfs contribute significantly to the total stellar mass of the SSP. In young stellar populations the light is dominated by the young stars that are still present in those populations and therefore the relative contribution of low-mass dwarfs to the integrated spectrum is even smaller. The situation is complicated further by the fact that the spectra of low-mass stars and K- and M-giants are very similar. However, there are a number of (gravity-sensitive) features in the spectrum that allow us to distinguish between dwarfs and giants (Wing & Ford 1969; Faber & French 1980; Gorgas et al. 1993; Worthey et al. 1994; Schiavon et al. 1997a,b; Cenarro et al. 2003; Schiavon 2007; Spiniello et al. 2012).

There is increasing evidence from spectroscopic studies of ETGs that the IMF is not universal. These studies suggest that ETGs with a higher mass/velocity dispersion/metallicity contain a higher fraction of low-mass stars (Conroy & van Dokkum 2012b; Spiniello et al. 2012; Ferreras et al. 2013; La Barbera et al. 2013; Martín-Navarro et al. 2015b). Within these galaxies, there appears to be a gradient that suggests that the IMF in the centre of the galaxy is more bottom-heavy than it is for more outer radii (Martín-Navarro et al. 2015a; van Dokkum et al. 2016). Combined studies of population synthesis, dynamics and/or gravitational lensing also find higher mass-to-light ratios for these galaxies than the MW-value (Treu et al.
Chapter 1. Introduction

2010; Graves & Faber 2010; Cappellari et al. 2012; Conroy & van Dokkum 2012b; Lyubenova et al. 2016). In contrast to these findings, Smith et al. (2015) and Newman et al. (2016) have found a number of nearby ETGs that are consistent with the MW IMF but the spread in slopes is still large.

In many astrophysical studies the IMF is a fundamental input quantity that is often assumed to be universal and similar to the MW IMF. For example, mass-to-light ratios of galaxies, the chemical evolution of galaxies, the energy balance of the interstellar medium, the population of stellar remnants, and the supernovae rate are all related to the IMF. If the IMF is not universal, this may affect any of the galaxy properties that are derived under the assumption of a universal IMF (Ferré-Mateu et al. 2013; Fontanot et al. 2017). Accurately measuring the shape of the IMF and possible variations of the IMF shape with other galaxy properties is therefore of crucial importance.

If one finds a correlation between for example the velocity dispersion of a galaxy and the slope of the IMF, this does not imply that there is a causal relation between these properties. In fact, it is highly unlikely that variations of the IMF are driven by global properties of the galaxy. Instead, one would expect that variations of the IMF are related to differences in the local properties of the clouds in which stars form, such as the pressure or the Mach number (Hennebelle & Chabrier 2008; Krumholz 2011; Hopkins 2012, 2013). These local properties may in turn reflect different typical circumstances of star formation in different galaxies which would then explain the empirical relation between for example IMF slope and velocity dispersion. Star formation is, however, a complex physical process that involves an interplay between gravity, turbulence, radiation, magnetic fields and the chemical composition of the gas. It is challenging to provide a theoretical framework work for this and the origin of the IMF in the process of star formation is not yet completely understood (Krumholz 2014; Offner et al. 2014). In that context, an accurate determination of how the IMF varies with other galaxy properties provides important constraints for any theory of star formation.

1.5 Bayesian model comparison

In the previous sections we have described how to construct a model for the spectrum of a stellar population, and the role of the IMF in the construction of such a model. Yet the ultimate goal is to extract the IMF from an
observed spectrum and to select the model that best fits the data. A statistically sound way to do this is through Bayesian model comparison.

Bayesian probability theory is based on two rules (Cox 1946) that may be used for manipulating conditional probabilities:

\[ p(H|I) + p(\bar{H}|I) = 1 \]  \hspace{1cm} (1.7)

and

\[ p(H, D|I) = p(D|H, I)p(H|I) = p(H|D, I)p(D|I) \]  \hspace{1cm} (1.8)

where \( H \) denotes the hypothesis, \( D \) denotes the data, \( \bar{H} \) stands for not \( H \), \( I \) represents any relevant background information and | should be read as “given that”. The first rule states that we are dealing with exclusive probabilities: adding the probability of \( H \) and the probability of not \( H \) is equal to one, where one means “true”. The second rule is the product rule and may be used to derive Bayes’ theorem:

\[ p(H|D, I) = \frac{p(D|H, I)p(H|I)}{p(D|I)} \]  \hspace{1cm} (1.9)

where \( p(H|D, I) \) represents the posterior probability, \( p(D|H, I) \) the likelihood function, \( p(H|I) \) the prior and \( p(D|I) \) is usually referred to as the evidence. The strength of Bayes’ theorem lies in its ability of transforming the likelihood, which is usually easier to calculate, into a posterior probability, which is the term we are actually interested in.

An important concept in Bayesian analysis is that of marginalization. The marginalization equation is given by

\[ p(x|I) = \int_{-\infty}^{+\infty} p(x, y|I)dy = \int_{-\infty}^{+\infty} p(x|y, I)p(y|I)dy. \]  \hspace{1cm} (1.10)

Marginalization may be used to include uncertainties from a nuisance model parameter \( y \) by integrating the conditional probability \( p(x|y, I) \) multiplied by the probability distribution of the nuisance parameter \( y \) over the complete parameter space of \( y \) to obtain the probability distribution of the parameter we are interested in (\( x \) in this case).

In the process of data modelling, we can distinguish at least two levels of inference. The first level of inference assumes that a given model \( M_0 \) is true and then fits the free parameters of that model to the data. We refer to the
Chapter 1. Introduction

first level of inference as parameter estimation. When Bayesian methods are used for parameter estimation, the results are in general very similar to those derived with classical statistical methods. The real difference between classical statistical methods and Bayesian methods is found in the second level of inference, where different models are compared in the light of the probability of the data (i.e. the evidence).

Models that are more complex can nearly always fit the data better. Therefore, comparing different models on the basis of the maximum likelihood one will nearly always select the more complex model. However, this more complex model is potentially over-parameterized. For example, consider the data in Figure 1.3 that we fit with two different models. The first model $M_0$ is a straight line with two free parameters, the second model $M_1$ a tenth-order polynomial with eleven free parameters. Although the second model provides a better fit to the data, the chances are that you think the first model is more appealing. Bayesian model comparison allows us to quantify the comparison between different models in the light of the data. Hence, Bayesian model comparison automatically embodies Occam’s razor, stating that we should select the “simplest” model that fits the data.

Suppose that we want to compare two models $M_0$ and $M_1$ to describe a given set of data $D$. The free parameters in the model are $\theta_0$ and $\theta_1$. At the first level of inference, we infer the posterior probability of the model parameters $\theta_i$. According to Bayes’ theorem, this probability distribution can be written as

$$ p(\theta_i|D, M_i) = \frac{p(D|\theta_i, M_i)p(\theta_i|M_i)}{p(D|M_i)}. $$

(1.11)

For the first level of inference the normalizing constant, or the evidence, is not important but for the second level of inference this term is very important.

When we compare the models $M_0$ and $M_1$, we once again apply Bayes’ theorem to calculate the odds ratio

$$ \frac{p(M_0|D)}{p(M_1|D)} = \frac{p(D|M_0)p(M_0)}{p(D|M_1)p(M_1)}, $$

(1.12)

where the normalizing constants cancel out in the numerator and denominator. Assuming that we assign equal prior probabilities to the two models, we can compare the models on the basis of the ratio $p(D|M_0)/p(D|M_1)$ which are the normalizing constants of Equation 1.11. In other words, to
1.5. Bayesian model comparison

Figure 1.3 – More complex models always fit the data better. According to Occam’s razor one should select the simplest model that fits the data.

compare models $M_0$ and $M_1$ we need to evaluate the evidence $p(D|M_0)$ for model $M_0$ and compare it to the evidence $p(D|M_1)$ for model $M_1$.

To understand how Bayesian model comparison automatically embodies Occam’s razor it is insightful to look at the simple case of a model $M$ with one free parameter $\theta$ (see also MacKay 1992, for a more extensive discussion). The evidence for that model may be written as

$$p(D|M) = \int p(D, \theta|M) d\theta = \int p(D|\theta, M)p(\theta|M)d\theta. \quad (1.13)$$

The first term in the integral is the likelihood of the data given the model and its free parameters, i.e., how well does the model fit the data. The second term is the prior for the probability distribution function of $\theta$. If
we assume that $\theta$ lies between $\theta_{\text{min}}$ and $\theta_{\text{max}}$ with a uniform prior, we may write

$$p(\theta|M) = \frac{1}{\theta_{\text{max}} - \theta_{\text{min}}} \text{ for } \theta_{\text{min}} \leq \theta \leq \theta_{\text{max}},$$  

(1.14)

and we have

$$p(D|M) = \frac{1}{\theta_{\text{max}} - \theta_{\text{min}}} \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} p(D|\theta, M)d\theta.$$  

(1.15)

Assuming that $p(D|\theta, M)$ can be described by a Gaussian centred at $\theta_0$ with standard deviation $\sigma_\theta$ that lies well within the prior range of $\theta$, the integral evaluates to $\sqrt{2\pi}\sigma_\theta$ so that

$$p(D|M) = p(D|\theta_0, M) \times \frac{\sqrt{2\pi}\sigma_\theta}{\theta_{\text{max}} - \theta_{\text{min}}} = L_{\text{max}}W.$$  

(1.16)

The second factor in this equation is referred to as the Occam factor and penalizes the model $M$ for having parameter $\theta$. Equation 1.16 shows that the evidence can be written as the product of the maximum likelihood $L_{\text{max}}$ times the Occam factor $W$. For models with one parameter, the Occam factor is the ratio of the width of the likelihood over the width of the prior. When the model includes multiple parameters this may be generalized to the ratio of the prior volume that provides a good fit to the data by the total volume that is accessible by the prior. For models with a broader prior or more parameters, the accessible prior volume is larger and hence the Occam factor becomes smaller and more complex models are automatically penalized. The evidence is therefore always a trade-off between the quality of the fit to the data and the model complexity. As long as an extra parameter improves the fit to the data sufficiently, the decrease of the Occam factor is compensated by an increase of the likelihood. However, if the likelihood only increases by a small amount when an additional parameter is included, Bayesian model comparison will prefer the simpler model with fewer parameters, because it has a higher probability to produce the data.
1.6 This thesis

The IMF is a fundamental quantity in the study of galaxy formation and evolution. A non-universal IMF has important consequences for many astrophysical studies and provides constraints for theories of star formation. Although there is increasing evidence that the IMF is not universal, the exact shape of the IMF and the scale of these IMF variations remain uncertain.

Inferring the IMF from the spectrum of a galaxy with a population synthesis model is not straightforward. Nowadays there are many different SPS models with a variety of different ingredients to choose from. Possible degeneracies between these ingredients and the inferred IMF as well as selecting which model one should use requires a solid statistical framework.

In this thesis we develop a hierarchical Bayesian framework for population synthesis that is specifically designed for inferring the IMF of unresolved stellar populations. However, the framework that we develop is more general and can be used to infer other properties of unresolved stellar populations as well. Within the model that we develop we use a parameterized IMF prior to regulate a more direct inference of the IMF shape. This direct inference gives more freedom to the model and allows the model to deviate from parameterized models when demanded by the data. The Bayesian framework of the model that we develop makes it well-suited for model comparison and allows us to objectively compare different ingredients of SPS models.

In Chapter 2 we discuss the general setup of the hierarchical Bayesian framework for SSPs. We combine the model with stellar templates that are based on the MILES library and use these templates to create a set of mock SSP spectra. Then we apply the model to these mock spectra and show that we can reconstruct the input parameters of the mock SSPs. However, when we apply the model to mock SSPs created with a different population synthesis code, this may introduce a bias on the inferred IMF parameters.

Many spectroscopic IMF studies of ETGs assume that an ETG can be modelled as an SSP. To test this assumption, in Chapter 3 we extend the SSP model of Chapter 2 to CSPs. We assume that we can model a CSP as a combination of SSPs and use the Bayesian evidence to discriminate between models with different numbers of SSPs. Using SFHs based on semi-analytic models, we create a number of realistic CSP mock spectra with an IMF that varies as a function of the velocity dispersion of the galaxy. Then we try
to reconstruct the IMF of these mock CSPs by using a variable number of SSPs in the fit.

One other motivation for starting the research described in this thesis has been the development of the X-shooter Spectral Library (XSL). The stars in XSL have been selected to provide a good coverage of the HR diagram and the spectra in the library extend all the way from the UV to the NIR. This makes XSL an ideal library for population synthesis studies of the IMF.

In Chapter 4, we combine the VIS arm of XSL with the MILES library to create the MIX stellar population models. Moreover, we extend the model developed in Chapter 2 and 3 to include various response functions to account for variations in the abundance pattern. Then, we combine the hierarchical Bayesian framework that we developed with the MIX stellar population models and apply this to a set of stacked SDSS spectra binned by velocity dispersion. When we fit these SDSS spectra, we include multiple SSPs, response functions of various elements, two different isochrone sets, two different regularization schemes and two different parameterizations of the IMF prior, and we compare different model ingredients on the basis of the Bayesian evidence.
Cox, R. 1946, Am. J. Phys., 14, 1
de Vaucouleurs, G. 1948, Annales d’Astrophysique, 11, 247
Hubble, E. P. 1936, Realm of the Nebulae
Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1
Liddle, A. 2003, An Introduction to Modern Cosmology (Wiley)
MacKay, D. J. C. 1992, Neural Computation, 4, 415
Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41