dessn Documentation

Release 0.0.1

dessn

March 06, 2016

CONTENTS

1	dessn	packa	ge	1	
	1.1	Subpa	ckages	1	
		1.1.1	dessn.entry package	1	
			Submodules	1	
			dessn.entry.sim module	1	
		1.1.2	dessn.model package	1	
			Submodules	1	
			dessn.model.edge module	1	
			dessn.model.model module	2	
			dessn.model.node module	3	
		1.1.3	dessn.simple package	6	
			Subpackages	6	
			Submodules	7	
			dessn.simple.example module	7	
			dessn.simple.exampleIntegral module	9	
			dessn.simple.exampleLatent module	10	
		1.1.4	dessn.simulation package	12	
			Submodules	12	
			dessn.simulation.observationFactory module	12	
			dessn.simulation.simulation module	12	
		1.1.5	dessn.utility package	13	
			Submodules	13	
			dessn.utility.hdemcee module	13	
			dessn.utility.newtonian module	13	
Pyt	Python Module Index 1				
Ind	ex			17	

CHAPTER

ONE

DESSN PACKAGE

Welcome to the **DESSN** code base.

1.1 Subpackages

1.1.1 dessn.entry package

Submodules

dessn.entry.sim module

1.1.2 dessn.model package

Submodules

dessn.model.edge module

An edge connection one or more parameters to one or more different parameters.

An edge is a connection between parameters (not Nodes), and signifies a joint probability distribution. That is, if in our mathematical definition of our model, we find the term P(a,b|c,d,e), this would be represented by a single edge. Similarly, P(a|b)P(b|c,d) would be two edges.

Parameters probability_of: str or list[str]

```
The dependent parameters. With the example P(a,b|c,d), this input would be ['a','b'].
```

given: str or list[str]

In the example P(a,b|c,d), this input would be ['c','d'].

get_log_likelihood(data)

Gets the log likelihood of this edge.

For example, if we had

$$P(a,b|c,d) = \frac{1}{\sqrt{2\pi}d} \exp\left(-\frac{(ab-c)^2}{d^2}\right),\,$$

we could implement this function as return -np.log(np.sqrt(2*np.pi)*data['d']) (data['a']*data['b'] - data['c'])**2/(data['d']**2)

Returns float

the log likelihood given the supplied data and the model parametrisation.

```
class dessn.model.edge.EdgeTransformation(inputs, transform_to)
```

Bases: dessn.model.edge.Edge

This specialised edge is used to connect to transformation nodes.

A transformation edge does not give a likelihood, but - as it is a known transformation - returns a dictionary when *get_transformation* is invoked that is injected into the data dictionary given to regular edges.

See LuminosityToAdjusted for a simple example.

Parameters inputs: str or list[str]

The required parameter inputs to do the transformation.

transform_to : str or list[str]

The parameters that will be added to the data dictionary after transformation

```
get_log_likelihood(data)
```

get_transformation(data)

Calculates the new parameters from the given data

Returns dict

a dictionary containing a value for each parameter given in transform_to

dessn.model.model module

```
{\bf class} \; {\tt dessn.model.model.Model} \; ({\it model\_name})
```

Bases: object

A generalised model for use in arbitrary situations.

A model is, at heart, simply a collection of nodes and edges. Apart from simply being a container in which to place nodes and edges, the model is also responsible for figuring out how to connect edges (which map to parameters) with the right nodes, for sorting edges such that when an edge is evaluated all its required data has been generated by other nodes or edges, for managing the emcee running, and also for generating the visual PGMs.

It is thus a complex class, and I expect, as of writing this summary, it contains numerous bugs.

Parameters model name: str

The model name, used for serialisation

add_edge (edge)

Adds an edge into the models collection of edges

add node (node)

Adds a node into the models collection of nodes.

finalise()

Finalises the model.

This method runs consistency checks on the model (making sure there are not orphaned nodes, edges to parameters that do not exist, etc), and in doing so links the right edges to the right nodes and determines the order in which edges should be evaluated.

fit_model (num_walkers=None, num_steps=5000, num_burn=3000, filename=None) Uses emcee to fit the supplied model.

This method sets an emcee run using the EnsembleSampler and manual chain management to allow for very high dimension models. MPI running is detected automatically for less hassle, and chain progress is serialised to disk automatically for convenience.

This method works... but is still a work in progress

Parameters num walkers: int, optional

The number of walkers to run. If not supplied, it defaults to eight times the model dimensionality

num_steps : int, optional

The number of steps to run

num_burn: int, optional

The number of steps to discard for burn in

filename: str, optional

If set, saves a corner plot to that filename in the top level plots directory.

Returns ndarray

```
The final flattened chain of dimensions (num_dimensions, num_walkers * (num_steps - num_burn))
```

get pgm(filename=None)

Renders (and returns) a PGM of the current model.

Parameters filename: str, optional

if the filename is set, the PGM is saved to file in the top level plots directory.

Returns daft.PGM

The daft PGM class, for further customisation if required.

dessn.model.node module

```
class dessn.model.node.Node (node_name, names, labels, parameter_type)
```

Bases: object

A node represented on a PGM model. Normally encapsulated by a single parameter, or several related parameters.

The Node class can essentially be thought of as a wrapper around a parameter or variable in your model. However, as some parameters are highly related (for example, flux and flux error), Nodes allow you to declare multiple parameters.

This class is an abstract class, and cannot be directly instantiated. Instead, instantiate one of the provided subclasses, as detailed below.

Parameters node_name : str

The node name, only used when plotting on a PGM

names : str or list[str]

The model parameter encapsulated by the node, or list of model parameters

labels: str or list[str]

Latex ready labels for the given names. Used in the PGM and corner plots.

parameter_type : NodeType

The type of subclass. Informs the model how to utilise the node.

class dessn.model.node.NodeLatent (node_name, names, labels)

Bases: dessn.model.node.Node

A node representing a latent, or hidden, variable in our model.

Given infinitely powerful computers, these nodes would not be necessary, for they represent marginalisation over unknown / kidden / latent parameters in the model, and we would simple integrate them out when computing the likelihood probability. However, this is not the case, and it is more efficient to simply incorporate latent parameters into our model and essentially marginalise over them using Monte Carlo integration. We thus trade explicit numerical integration in each step of our calculation for increased dimensionality.

For examples on why and how to use latent parameters, see the examples beginning in Example.

Parameters node_name : str

The node name, only used when plotting on a PGM

names: str or list[str]

The model parameter encapsulated by the node, or list of model parameters

labels: str or list[str]

Latex ready labels for the given names. Used in the PGM and corner plots.

get num latent()

The number of latent parameters to include in the model.

Running MCMC requires knowing the dimensionality of our model, which means knowing how many latent parameters (realisations of an underlying hidden distribution) we require.

For example, if we observe a hundred supernova drawn from an underlying supernova distribution, we would have to realise a hundred latent variables - one per data point.

Returns int

the number of latent parameters required by this node

class dessn.model.node.NodeObserved (node_name, names, labels, datas)

Bases: dessn.model.node.Node

A node representing one or more observed variables

This node is used for all observables in the model. In addition to a normal node, it also contains data, which can be in arbitrary format. This data is what is given to the incoming and outgoing node edges to calculate likelihoods.

Parameters node name: str

The node name, only used when plotting on a PGM

names : str or list[str]

The model parameter encapsulated by the node, or list of model parameters

labels: str or list[str]

Latex ready labels for the given names. Used in the PGM and corner plots.

datas : object or list[obj]

One data object for each supplied parameter name. **Must** be the same length as names if names is a list.

```
get_data()
```

Returns a dictionary containing keys of the parameter names and values of the parameter data object

class dessn.model.node.NodeTransformation(node_name, names, labels)

Bases: dessn.model.node.Node

A node representing a variable transformation.

This node essentially represents latent variables which are fully determined - their probability is given by a delta function. Examples of this might be the luminosity distance, as it is known exactly when given cosmology and redshift. Or it might represent a conversion between observed flux and actual flux, given we have a well defined flux correction.

On a PGM, this node would be represented by a point, not an ellipse.

Note that this node declares all associated parameters to be transformation parameters, although the transformation functions themselves are defined by the edges into and out of this node.

Parameters node name: str

The node name, only used when plotting on a PGM

names: str or list[str]

The model parameter encapsulated by the node, or list of model parameters

labels: str or list[str]

Latex ready labels for the given names. Used in the PGM and corner plots.

class dessn.model.node.NodeType

Bases: enum. Enum

LATENT = <NodeType.LATENT: 3>

OBSERVED = <NodeType.OBSERVED: 2>

TRANSFORMATION = <NodeType.TRANSFORMATION: 4>

UNDERLYING = <NodeType.UNDERLYING: 1>

class dessn.model.node.NodeUnderlying(node_name, names, labels)

Bases: dessn.model.node.Node

A node representing an underlying parameter in your model.

On the PGM, these nodes would be at the very top, and would represent the variables we are trying to fit for, such as Ω_M .

These nodes are required to implement the abstract method get_log_prior

Parameters node_name : str

The node name, only used when plotting on a PGM

names : str or list[str]

The model parameter encapsulated by the node, or list of model parameters

labels: str or list[str]

Latex ready labels for the given names. Used in the PGM and corner plots.

get_log_prior(data)

Returns the log prior for the parameter.

Parameters data: dic

A dictionary containing all data and the model parameters being tested at a given step in the MCMC chain. For this class, if the class was instantiated with a name of "omega_m", the input dictionary would have the key "omega_m", and the value of "omega_m" at that particular step in your chain.

Returns float

the log prior probability given the current value of the parameters

1.1.3 dessn.simple package

This module is designed to give a step by step overview of a very simplified example Bayesian model.

The basic example model is laid out in the parent class <code>Example</code>, and there are three implementations. The first implementation, <code>ExampleIntegral</code>, shows how the problem might be approached in a simple model, where numerical integration is simply done as part of the likelihood calculation.

However, if there are multiple latent parameters, we get polynomial growth of the number of numerical integrations we have to do, and so this does not scale well at all.

This leads us to the implementation in *ExampleLatent*, where we use the MCMC algorithm to essentially do Monte Carlo integration via marginalisation. This means we do not need to perform the numerical integration in the likelihood calculation, however the cost of doing so is increase dimensionality of our MCMC.

Finally, the <code>ExampleModel</code> implementation shows how the <code>ExampleLatent</code> class might be written to make use of Nodes. This is done in preparation for more complicated models, which will have more than one layer and needs to be configurable.

Subpackages

dessn.simple.modelbased package

I have placed the class based example for implementing the simplified model into its own module, so that the documentation generating for the simple module does not get cluttered with all the small classes this module will have.

The primary class to look at in code is the <code>ExampleModel</code> class.

I should finally note that in order to demonstrate parameter transformations, I have modified the model used in the previous two examples (<code>ExampleIntegral</code> and <code>ExampleLatent</code>) to also include a luminosity transformation, where I simply halve the luminosity before converting it to flux. Physically, this could represent a perfect 50% mirror absorption on the primary telescope mirror.

Submodules

$dessn. simple. model based. example Latent Class\ module$

 ${\bf class} \; {\tt dessn.simple.modelbased.exampleLatentClass.ExampleModel}$

Bases: dessn.model.model.Model

An implementation of ExampleLatent using classes instead of procedural code.

The model is set up by declaring nodes, the edges between nodes, and then calling finalise on the model to verify its correctness.

This is the primary class in this package, and you can see that other classes inherit from either Node or from Edge.

I leave the documentation for Node's and :class:.Edge's to those classes, and encourage viewing the code directly to understand exactly what is happening.

class dessn.simple.modelbased.exampleLatentClass.FluxToLuminosity

Bases: dessn.model.edge.Edge

get_log_likelihood(data)

class dessn.simple.modelbased.exampleLatentClass.LatentLuminosity (n=100)

Bases: dessn.model.node.NodeLatent

get_num_latent()

class dessn.simple.modelbased.exampleLatentClass.LuminosityToAdjusted

Bases: dessn.model.edge.EdgeTransformation

get_transformation(data)

 ${\bf class}~{\tt dessn.simple.modelbased.exampleLatentClass.LuminosityToSupernovaDistribution}$

Bases: dessn.model.edge.Edge

get_log_likelihood(data)

 ${f class}$ dessn.simple.modelbased.exampleLatentClass. ${f ObservedFlux}$ ($n{=}100$)

Bases: dessn.model.node.NodeObserved

class dessn.simple.modelbased.exampleLatentClass.UnderlyingSupernovaDistribution

Bases: dessn.model.node.NodeUnderlying

get_log_prior(data)

class dessn.simple.modelbased.exampleLatentClass.UselessTransformation

Bases: dessn.model.node.NodeTransformation

Submodules

dessn.simple.example module

class dessn.simple.example.Example (n=30, theta 1=100.0, theta 2=20.0)

Bases: object

Setting up the math for some examples.

Let us assume that we are observing supernova that are drawn from an underlying supernova distribution parameterised by θ , where the supernova itself simply a luminosity L. We measure the luminosity of multiple supernovas, giving us an array of measurements D. If we wish to recover the underlying distribution of supernovas from our measurements, we wish to find $P(\theta|D)$, which is given by

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

Note that in the above equation, we realise that $P(D|L) = \prod_{i=1}^{N} P(D_i|L_i)$ as our measurements are independent. The likelihood $P(D|\theta)$ is given by

$$P(D|\theta) = \prod_{i=1}^{N} \int_{-\infty}^{\infty} P(D_i|L_i) P(L_i|\theta) dL_i$$

We now have two distributions to characterise. Let us assume both are gaussian, that is our observed luminosity x_i has gaussian error σ_i from the actual supernova luminosity, and the supernova luminosity is drawn from an underlying gaussian distribution parameterised by θ .

$$P(D_i|L_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - L_i)^2}{\sigma_i^2}\right)$$
$$P(L_i|\theta) = \frac{1}{\sqrt{2\pi}\theta_2} \exp\left(-\frac{(L_i - \theta_1)^2}{\theta_2^2}\right)$$

This gives us a likelihood of

$$P(D|\theta) = \prod_{i=1}^{N} \frac{1}{2\pi\theta_2 \sigma_i} \int_{-\infty}^{\infty} \exp\left(-\frac{(x_i - L_i)^2}{\sigma_i^2} - \frac{(L_i - \theta_1)^2}{\theta_2^2}\right) dL_i$$

Working in log space for as much as possible will assist in numerical precision, so we can rewrite this as

$$\log (P(D|\theta)) = \sum_{i=1}^{N} \left[\log \left(\int_{-\infty}^{\infty} \exp \left(-\frac{(x_i - L_i)^2}{\sigma_i^2} - \frac{(L_i - \theta_1)^2}{\theta_2^2} \right) dL_i \right) - \log(2\pi\theta_2\sigma_i) \right]$$

Parameters n: int, optional

The number of supernova to 'observe'

theta_1: float, optional

The mean of the underlying supernova luminosity distribution

theta_2: float, optional

The standard deviation of the underlying supernova luminosity distribution

do_emcee (nwalkers=None, nburn=None, nsteps=None)

Abstract method to configure the emcee parameters

get_likelihood(theta, data, error)

Abstract method to return the log likelihood

get_posterior (theta, data, error)

Gives the log posterior probability given the supplied input parameters.

Parameters theta: array of model parameters

 \mathbf{data} : array of length n

An array of observed luminosities

error : array of length *n*

An array of observed luminosity errors

Returns float

the log posterior probability

get_prior (theta)

Get the log prior probability given the input.

The prior distribution is currently implemented as flat prior.

Parameters theta: array of model parameters

Returns float

the log prior probability

plot_observations()

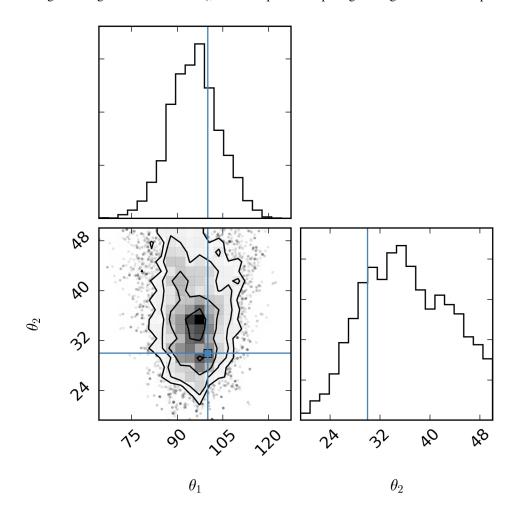
Plot the observations and observation distribution.

dessn.simple.exampleIntegral module

class dessn.simple.exampleIntegral.ExampleIntegral(n=10, theta_1=100.0, theta_2=30.0)
 Bases: dessn.simple.example.Example

An example implementation using integration over a latent parameter.

Building off the math from Example Creating this class will set up observations from an underlying distribution. Invoke emcee by calling the object. In this example, we perform the marginalisation inside the likelihood calculation, which gives us dimensionality only of two (the length of the θ array). However, this is at the expense of performing the marginalisation over dL_i , as this requires computing n integrals for each step in the MCMC.



Parameters n: int, optional

The number of supernova to 'observe'

theta_1: float, optional

The mean of the underlying supernova luminosity distribution

theta_2: float, optional

The standard deviation of the underlying supernova luminosity distribution

 $\verb"do_emcee" (nwalkers=20, nburn=2500, nsteps=3000)$

Run the *emcee* chain and produce a corner plot.

Saves a png image of the corner plot to plots/exampleIntegration.png.

Parameters nwalkers: int, optional

The number of walkers to use. Minimum of four.

nburn: int, optional

The burn in period of the chains.

nsteps: int, optional

The number of steps to run

get_likelihood(theta, data, error)

Gets the log likelihood given the supplied input parameters.

Parameters theta: array of size 2

An array representing $[\theta_1, \theta_2]$

 \mathbf{data} : array of length n

An array of observed luminosities

error: array of length n

An array of observed luminosity errors

Returns float

the log likelihood probability

dessn.simple.exampleLatent module

class dessn.simple.exampleLatent.ExampleLatent (n=30, theta_1=100.0, theta_2=20.0)

Bases: dessn.simple.example.Example

An example implementation using marginalisation over latent parameters.

Building off the math from Example, instead of performing the integration numerically in the computation of the likelihood, we can instead use Monte Carlo integration by simply setting the latent parameters \vec{L} as free parameters, giving us

$$\log \left(P(D|\theta, \vec{L}) \right) = -\sum_{i=1}^{N} \left[\frac{(x_i - L_i)^2}{\sigma_i^2} + \frac{(L_i - \theta_1)^2}{\theta_2^2} + \log(2\pi\theta_2\sigma_i) \right]$$

Creating this class will set up observations from an underlying distribution. Invoke emcee by calling the object. In this example, we marginalise over L_i after running our MCMC, and so we no longer have to compute integrals in our chain, but instead have dimensionality of 2 + n, where n are the number of observations.

Parameters n: int, optional

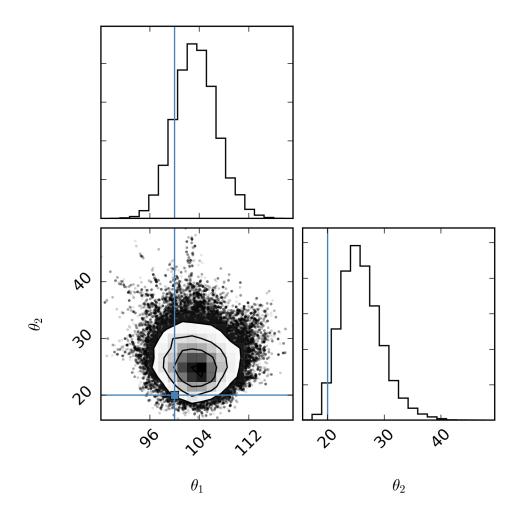
The number of supernova to 'observe'

theta_1: float, optional

The mean of the underlying supernova luminosity distribution

theta_2: float, optional

The standard deviation of the underlying supernova luminosity distribution



```
do_emcee (nwalkers=500, nburn=2000, nsteps=2500)
          Run the emcee chain and produce a corner plot.
          Saves a png image of the corner plot to plots/exampleLatent.png.
               Parameters nwalkers: int, optional
                     The number of walkers to use.
                   nburn: int, optional
                     The burn in period of the chains.
                   nsteps: int, optional
                     The number of steps to run
     get_likelihood(theta, data, error)
          Gets the log likelihood given the supplied input parameters.
               Parameters theta: array of length 2 + n
                     An array representing [\theta_1, \theta_2, \vec{L}]
                   \mathbf{data}: array of length n
                     An array of observed luminosities
                  {\bf error} : array of length \boldsymbol{n}
                     An array of observed luminosity errors
               Returns float
                     the log likelihood probability
1.1.4 dessn.simulation package
Submodules
dessn.simulation.observationFactory module
class dessn.simulation.observationFactory.ObservationFactory(**kwargs)
     Bases: object
     check_kwargs()
     get observations(num)
          Still needs massive refactoring
dessn.simulation.simulation module
class dessn.simulation.simulation.Simulation
     Bases: object
     get_simulation(num_trans=30)
```

1.1.5 dessn.utility package

Submodules

dessn.utility.hdemcee module

dessn.utility.newtonian module

d

```
dessn, 1
dessn.entry, 1
dessn.entry.sim, 1
dessn.model, 1
dessn.model.edge, 1
dessn.model.model, 2
dessn.model.node, 3
dessn.simple, 6
dessn.simple.example,7
dessn.simple.exampleIntegral, 9
dessn.simple.exampleLatent, 10
dessn.simple.modelbased, 6
dessn.simple.modelbased.exampleLatentClass,
       6
dessn.simulation, 12
dessn.simulation.observationFactory, 12
dessn.simulation.simulation, 12
dessn.utility, 13
dessn.utility.hdemcee, 13
dessn.utility.newtonian, 13
```

16 Python Module Index

A add_edge() (dessn.model.model.Model method), 2 add_node() (dessn.model.model.Model method), 2 C check_kwargs() (dessn.simulation.observationFactory.Observation), 12 D	ExampleLatent (class in dessn.simple.exampleLatent), 10 ExampleModel (class in dessn.simple.modelbased.exampleLatentClass), 6 ForvationFactory finalise() (dessn.model.model.Model method), 2 fit() (dessn.utility.newtonian.NewtonianPosition method), 13
dessn (module), 1 dessn.entry (module), 1 dessn.entry.sim (module), 1 dessn.model (module), 1	fit_model() (dessn.model.model.Model method), 2 FluxToLuminosity (class in dessn.simple.modelbased.exampleLatentClass), 7
dessn.model.edge (module), 1 dessn.model.model (module), 2 dessn.simple (module), 6 dessn.simple (module), 6 dessn.simple.example (module), 7 dessn.simple.exampleIntegral (module), 9 dessn.simple.exampleLatent (module), 10 dessn.simple.modelbased (module), 6 dessn.simple.modelbased.exampleLatentClass (module), 6 dessn.simulation (module), 12 dessn.simulation.observationFactory (module), 12 dessn.simulation.simulation (module), 12 dessn.utility (module), 13 dessn.utility.hdemcee (module), 13 dessn.utility.newtonian (module), 13 do_emcee() (dessn.simple.example.Example method), 8 do_emcee() (dessn.simple.exampleIntegral.ExampleIntegral method), 9 do_emcee() (dessn.simple.exampleLatent.ExampleLatent method), 10	get_data() (dessn.model.node.NodeObserved method), 5 get_likelihood()
Edge (class in dessn.model.edge), 1 EdgeTransformation (class in dessn.model.edge), 2 EmceeWrapper (class in dessn.utility.hdemcee), 13 Example (class in dessn.simple.example), 7 ExampleIntegral (class in dessn.simple.exampleIntegral), 9	get_num_latent() (dessn.simple.modelbased.exampleLatentClass.LatentLurmethod), 7 get_observations() (dessn.simulation.observationFactory.ObservationFactor method), 12 get_pgm() (dessn.model.model.Model method), 3 get_posterior() (dessn.simple.example.Example method),

method), 13

```
S
                 (dessn.utility.hdemcee.EmceeWrapper
get results()
         method), 13
                                                      Simulation (class in dessn.simulation.simulation), 12
get simulation() (dessn.simulation.simulation.Simulation
         method), 12
                                                      Τ
get_transformation() (dessn.model.edge.EdgeTransformation_TRANSFORMATION (dessn.model.node.NodeType at-
         method), 2
method), 7
                                                     UNDERLYING (dessn.model.node.NodeType attribute),
            (dessn.utility.newtonian.NewtonianPosition
iterate()
                                                     UnderlyingSupernovaDistribution
         method), 13
                                                               dessn.simple.modelbased.exampleLatentClass),
                                                      UselessTransformation
                                                                                      (class
                                                                                                       in
LATENT (dessn.model.node.NodeType attribute), 5
                                                               dessn.simple.modelbased.exampleLatentClass),
LatentLuminosity
                              (class
                                                  in
         dessn.simple.modelbased.exampleLatentClass),
LuminosityToAdjusted
                                (class
                                                  in
         dessn.simple.modelbased.exampleLatentClass),
LuminosityToSupernovaDistribution
         dessn.simple.modelbased.exampleLatentClass),
M
Model (class in dessn.model.model), 2
Ν
NewtonianPosition (class in dessn.utility.newtonian), 13
Node (class in dessn.model.node), 3
NodeLatent (class in dessn.model.node), 4
NodeObserved (class in dessn.model.node), 4
NodeTransformation (class in dessn.model.node), 5
NodeType (class in dessn.model.node), 5
NodeUnderlying (class in dessn.model.node), 5
\mathbf{O}
ObservationFactory
                               (class
                                                 in
         dessn.simulation.observationFactory), 12
OBSERVED (dessn.model.node.NodeType attribute), 5
ObservedFlux
         dessn.simple.modelbased.exampleLatentClass),
Р
plot()
            (dessn.utility.newtonian.NewtonianPosition
         method), 13
plot observations()
                      (dessn.simple.example.Example
         method), 8
R
run_chain()
                 (dessn.utility.hdemcee.EmceeWrapper
```

18 Index