Tutorial on Bayesian Inference (2020)

1) Extragalactic radio sources

Imagine a population of extragalactic radio sources: these are complicated objects which exhibit intrinsic variations. This introduces randomness. The distribution of their flux densities is well described by a power law with slope $-\alpha$. The likelihood to measure a radio flux S is then

$$\mathcal{P}(S|\alpha) \propto S^{-\alpha}. \tag{1}$$

1. This distribution diverges at $S = 0$, but telescopes can only detect a minimal limiting flux S_0. For $S_0 > 0$, normalize the distribution.
2. $\mathcal{P}(S|\alpha)$ describes how frequently you will find a source with flux S if the Universe uses some value for α. Imagine now the inverse problem: You measured multiple fluxes S and now want to infer α. Which distribution do you have to set up? Using a uniform prior and Bayes theorem, derive this distribution.
3. Imagine a single source has been observed, and its flux has the value $2S_0$. Which calculation do you have to carry out in order to infer the most likely value of α after this single observation? Show that the most likely value is $\alpha = 2.44$.
4. Plot the posterior of α. If you quoted a standard deviation for the uncertainty of α, which information would this omit?

Historical sidenote: For a non-evolving Euclidean Universe, one finds $\alpha = 3/2$. In the 1960s, α was detected to not be $3/2$, and this finding was interpreted as evidence against the then prevalent steady-state cosmology.

2) Modelling and inferring stellar properties

According to the Stefan-Boltzmann law, the luminosity L of a star (seen as a black body) scales with its area A and temperature T as

$$L = \sigma_T A T^4. \tag{2}$$

1. Stellar temperatures result from a complicated stellar formation process and therefore vary between the stars. Assume here that the stellar temperatures are drawn from

   $$T \sim \mathcal{G}(T_0, \sigma). \tag{3}$$

   Derive the distribution $\mathcal{P}(L)$ of the resulting stellar luminosities.
2. For $T_0 = 10, \sigma = 1, A = 1$, plot the distribution function of stellar luminosities.
3. Generate random samples from $\mathcal{P}(L)$, histogram them, and show that the histogram approximates $\mathcal{P}(L)$ for sufficiently many samples.

3) Satellites: the next generation

Measurement campaigns are frequently organized in stages. For example, an elderly satellite will be replaced by a new satellite with the same scientific task but better design. This happened e.g. when WMAP was followed up by Planck. In this context, the posterior of an old experiment can be reinterpreted as a 'datadriven prior', and can then be combined with the likelihood of the new experiment, to yield the updated posterior. Here, we investigate this interplay of priors and likelihoods.
Assume a data point y has a Gaussian sampling distribution with mean θ and variance σ^2

$$y \sim \mathcal{G}(\theta, \sigma^2). \tag{4}$$

The Gaussian is self-conjugate, i.e. a Gaussian prior times a Gaussian likelihood will lead to a Gaussian posterior. Therefore, let us use the prior

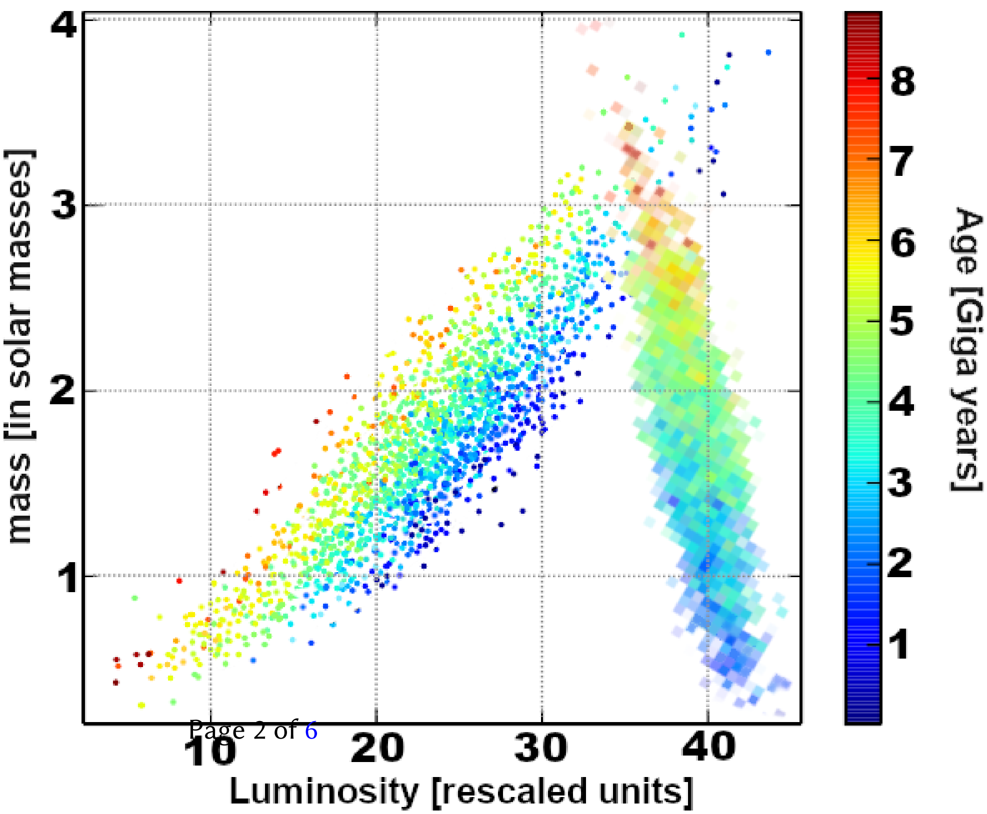
$$\theta \sim \mathcal{G}(\mu_0, \tau_0^2). \tag{5}$$

1. Which hyper-parameters occur here?
2. Derive the posterior $\mathcal{P}(\theta|y)$.
3. Plot prior and posterior.
4. For which values of the hyper-parameters do you have a highly informative, or uninformative prior?
5. If the prior describes the older experiment, what does an informative prior then express?
6. If the prior does not describe any older experiment, but simply an assumed distribution which one needs to pick because one needs a prior after all... what does the informative prior then describe?
7. How do you have to set the hyper-parameters to reach a limit where the prior disagrees with the sampling distribution on likely values for θ? If the prior describes an older experiment, what does this disagreement then describe?

4) Astronomical objects: Conditionals versus Marginals

Consider Fig. 1 which displays intrinsic variabilities of two stellar populations. Population A is indicated as dots, population B is indicated via pixels. The axes display the objects' masses m, luminosities L and age a.

1. Sketch the marginals of stellar luminosities for the two populations.
2. Sketch the marginals of stellar masses for the two populations.



3. Sketch the marginal of luminosities, independent of the populations.

4. Sketch the marginal of luminosities for objects from population A who have an age less than 3 giga years.

5. Sketch $\mathcal{P}(m|B, a > 6)$.

6. Sketch $\mathcal{P}(L|B, m = 2)$.

7. Sketch $\mathcal{P}(m|A, L = 15)$.

8. Sketch $\mathcal{P}(m|a > 7)$. What does this distribution describe?

The plot is repeated multiple times at the end of the problem sheet, such that you can draw on the plot itself. If you are colour blind, please signal this before the exam. The colour bar ranges from red (top), to yellow and green (middle), to blue (bottom).

5) Straight line fitting with a twist ....

Power law relations and linear relations are common in astronomical research. Power law relations turn into linear relations when a logarithmic quantity is considered. For example, the Faber-Jackson relation for early-type galaxies connects luminosity L and stellar velocity dispersion σ via

$$L \propto \sigma^{\gamma} \quad \Rightarrow \quad \log(L) \propto \gamma \log(\sigma). \tag{6}$$

If one wishes to infer γ, then magnitudes (logarithmic in L) of galaxies will come with an error, and so will the velocity dispersions. One hence faces the situation of wanting to infer a slope m of a straight line

$$y = mx \tag{7}$$

but we only have noisy estimates $\hat{x}_i$ and $\hat{y}_i$.
Assume there is just a single data point $(\hat{x}, \hat{y})$. Assume further that

$$\hat{x} \sim \mathcal{G}(x, \sigma^2), \text{ and independently } \hat{y} \sim \mathcal{G}(y, \sigma^2). \tag{8}$$

How do we then infer m?

1. Does the straight line with slope m relate the true values or the noisy values?

2. What is $\mathcal{P}(\hat{x}|x)$, what is $\mathcal{P}(\hat{y}|y)$?

3. What is $\mathcal{P}(\hat{x}, \hat{y}|x, y)$?

4. Give $\mathcal{P}(m|x, y)$ and $\mathcal{P}(y|x, m)$.

5. We are searching for the posterior $\mathcal{P}(m|\hat{x}, \hat{y})$. Use Bayes' theorem and a prior to invert the conditional dependence.

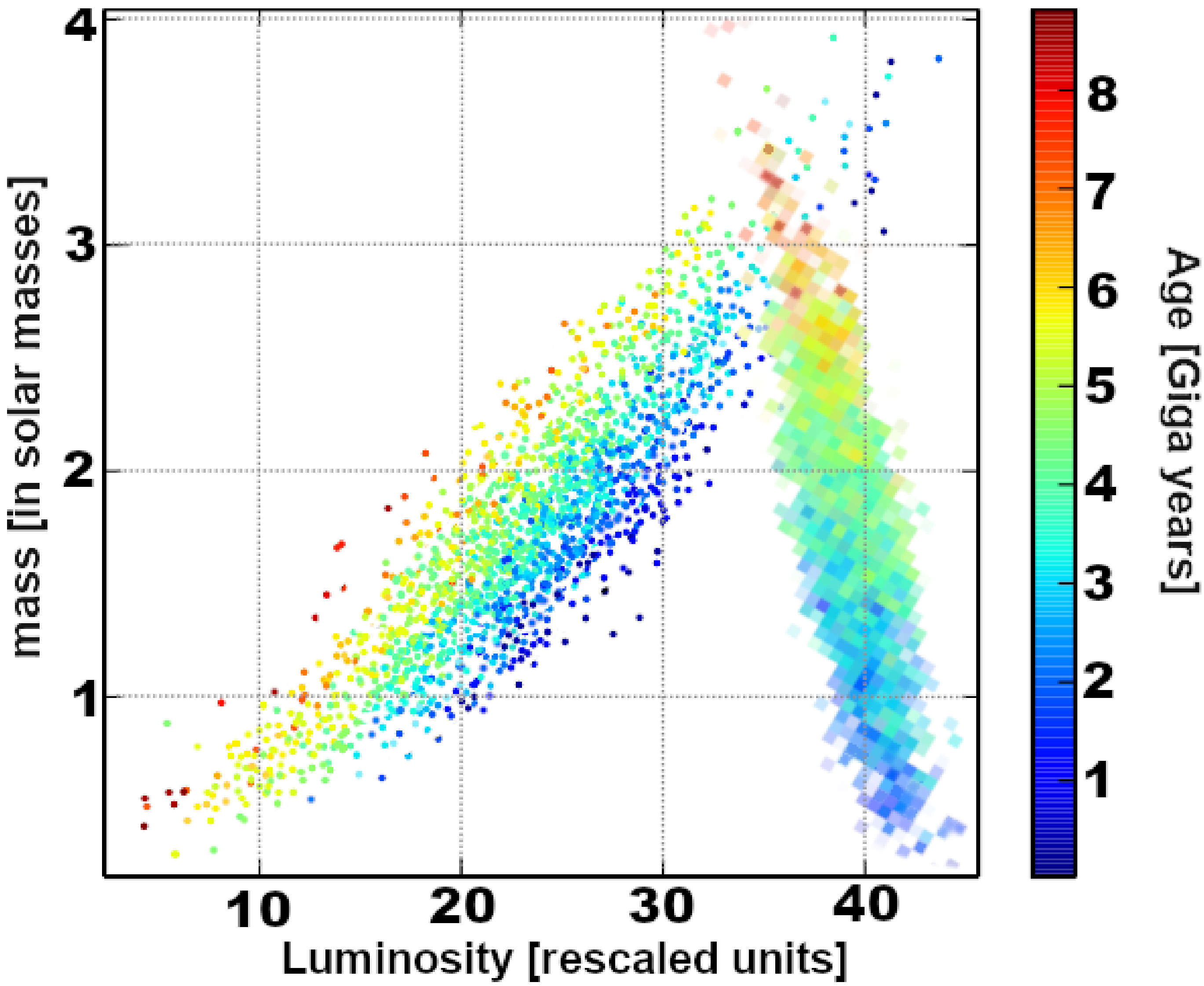
6. The above steps are preparations for simplyfying the Bayesian Hierarchical Model, which is the current step: To get $\mathcal{P}(m|\hat{x}, \hat{y})$ one always needs to start from the joint probability which includes all variables in the game. In our case, the starting point is therefore

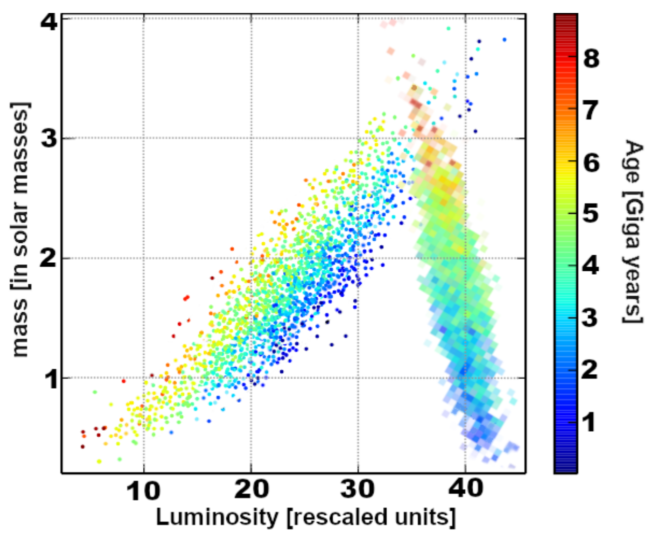
$$\mathcal{P}(\hat{x}, \hat{y}, x, y, m). \tag{9}$$

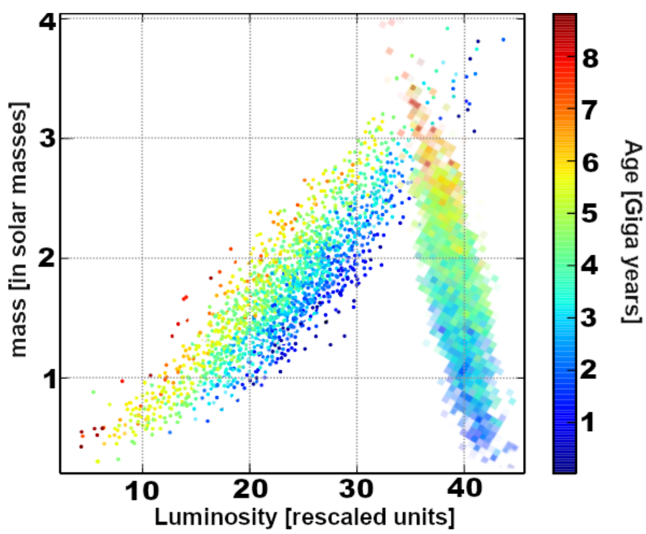
Manipulate Eq. (9) until you have $\mathcal{P}(m|\hat{x}, \hat{y})$ as a function of distributions which are all known. It is advisable to not yet plug in the Gaussians $\mathcal{G} = \exp(...)$, it would only complicate the situation. Staying on the level of $\mathcal{P}(\cdot|\cdot)$ is completely sufficient.

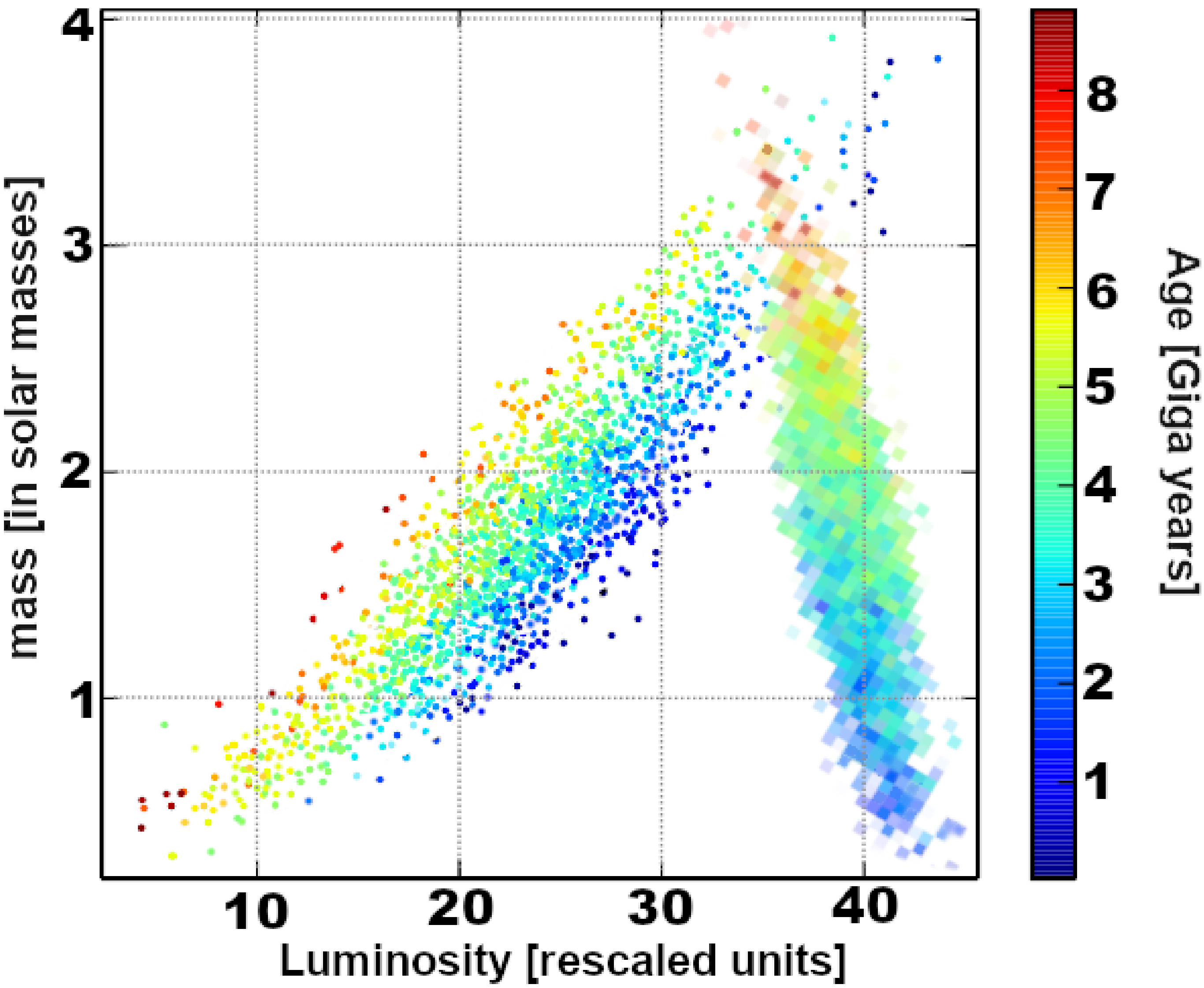
7. Using $\sigma = 1$, and uniform priors on x, y, m, now plug in the Gaussians and show that the posterior is

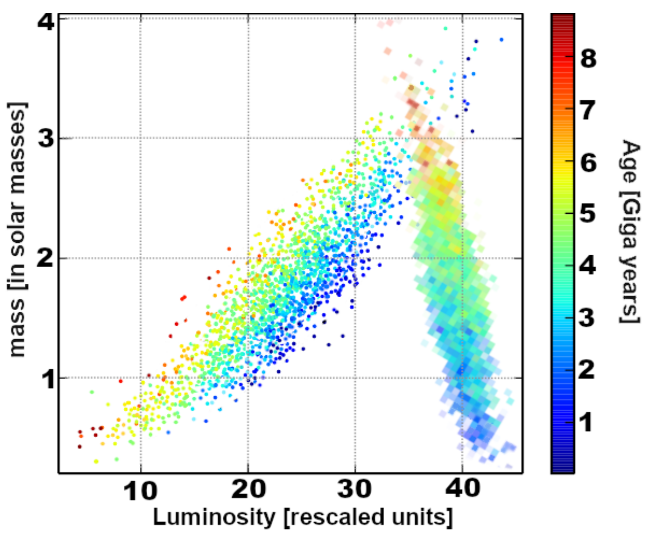
$$\mathcal{P}(m|\hat{x}, \hat{y}) \propto \frac{1}{1+m^2} \exp\left(-\frac{1}{2}\frac{(\hat{y} - m\hat{x})^2}{1+m^2}\right) \tag{10}$$

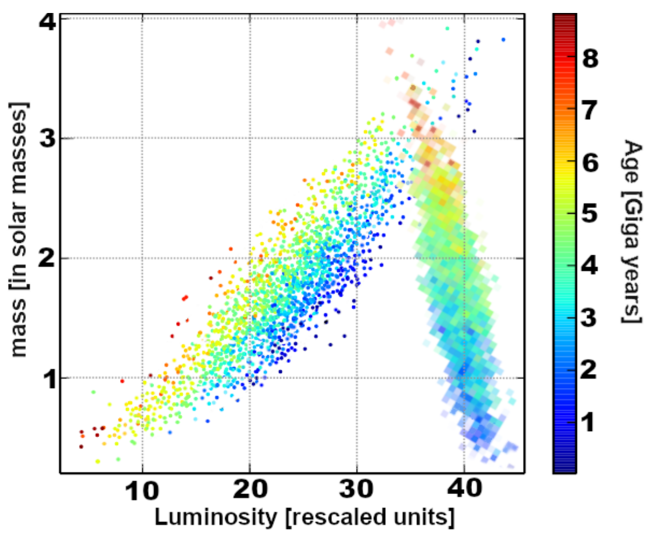

mass [in solar masses]
Luminosity [rescaled units]
Age [Giga years]
10
20
30
40
1
2
3
4
1
2
3
4
5
6
7
8


mass [in solar masses]
Luminosity [rescaled units]
Age [Giga years]


mass [in solar masses]
Luminosity [rescaled units]
Age [Giga years]


mass [in solar masses]
4
3
2
1
10
20
30
40
Luminosity [rescaled units]
8
7
6
5
4
3
2
1
Age [Giga years]


mass [in solar masses]
4
3
2
1
10
20
30
40
Luminosity [rescaled units]
8
7
6
5
4
3
2
1
Age [Giga years]


mass [in solar masses]
4
3
2
1
10
20
30
40
Luminosity [rescaled units]
8
7
6
5
4
3
2
1
Age [Giga years]