Notes on the program QSOSIM9 - John Webb, Dec 2013

June 11, 2014

This program generates simulated spectra including the forest and including up to 20 additional user-specified absorption systems (presumably DLAs or LLSs). It only uses HI at present. Writes normalised spectrum to a new ascii file, *spec.dat*, 4-column format: λ, flux, σ, no-noise-flux.

Random selection of redshifts for forest:
QSOSIM9 uses the simplest possible redshift evolution for the forest. Lines are unclustered and the number of lines per unit redshift is taken as

$$dn/dz = A(1+z)^{\gamma}$$

γ is a user-input via *sim.dat* (see below). A is the normalisation measured for a specific detection threshold (and is hard coded). Penton et al 2004 find $\gamma = 1.85$ for $z > 2.5$ and $\gamma = 0.16$ for $z < 2.5$, so note the approximation used in QSOSIM9. The default in *sim.dat* is $\gamma = 2$. Integrating to get the total number of lines detected in the range $z_1 - z_2$,

$$n = \frac{A}{\gamma+1}\left[(1+z_2)^{\gamma+1} - ((1+z_1)^{\gamma+1})\right]$$

The cumulative probability function is therefore

$$c = \frac{(1+z)^{\gamma+1} - ((1+z_1)^{\gamma+1}}{(1+z_2)^{\gamma+1} - ((1+z_1)^{\gamma+1}}$$

for $z_1 < z < z_2$, i.e. c is a uniform random variable, $0 < c < 1$.
Let $p = (1+z_2)^{\gamma+1}$, $q = (1+z_1)^{\gamma+1}$. Then

$$c(p-q) + q = (1+z)^{\gamma+1}$$

$$\left[\frac{\log_{10}(c(p-q)+q)}{\gamma+1}\right] = \log_{10}(1+z) = x$$

Re-arranging for z,

$$z = 10^x - 1$$

Random selection of N_{HI} for forest:
A single power law distribution is used, with lower cutoff. The number of lines per unit column density interval is

$$\frac{dn}{dN} \propto N^{-\beta}$$

where $\beta = 1.7$ is currently hard-coded. Note Penton et al 2004 suggest $\beta = 1.65 \pm 0.007$ for $10^{12.5} < N < 10^{14.5}$ and a flatter slope for $N > 10^{14.5}$.
Use the same procedure as above, i.e. integrate and form the cumulative probability function again,

$$d = \frac{N_c^{1-\beta} - N^{1-\beta}}{N_c^{1-\beta} - N_{hi}^{1-\beta}}$$

where d is a uniform random variable, $0 < d < 1$, N_c and N_{hi} are the low and high cut-offs. $N_c = 3e13$ is the default value in *sim.dat*.

To a good approximation (since $N_{hi}^{1-\beta} \sim 10^{22}$), this is

$$d = \frac{N_c^{1-\beta} - N^{1-\beta}}{N_c^{1-\beta}}$$

This re-arranges for N to give

$$\log_{10}(N) = \frac{\log_{10}(1-d)}{1-\beta} + \log_{10}(N_c)$$

Random selection of b for forest:
This is a straightforward Gaussian, with default mean and standard deviation 23 and 3 km/s.

Random number generators:
ran3 (Numerical Recipes) is used for uniform random numbers. Don't use *ran1* - it doesn't seem to work with gfortran on a Mac.

Error array:
There are 2 simple models. One produces a constant S/N per pixel in the continuum (no wavelength dependence) but uses a base-value at zero-flux levels (parameter inoise=0 in *sin.dat*). For this model, the 1-σ error array, σ, is hard-coded as

$$\sigma = \frac{I_{obs}}{s/n} + 0.2\frac{I_0}{s/n}$$

where s/n is the signal-to-noise parameter given in *sim.dat*, I_{obs} is the observed intensity, I_0 is the unabsorbed intensity, so note that the "actual" signal-to-noise is not equal to the number entered in *sim.dat*. The reason for doing this (and not simply using $\sqrt{n}$ statistics) is saturated lines and Lyman limits would remain noise-free.

The other model (parameter inoise=1 in *sin.dat*) degrades the noise towards the blue, using:

$$noise \propto (1 + e^{-x})$$

where

$$x = \frac{\lambda - c_1}{c_2}$$

and where default (hard-coded) values are $c_1 = 3532$ and $c_2 = 117$.

Input file *sin.dat*:

3.0	Quasar emission redshift.
-0.7	Spectral index.
16.0	V magnitude.
3000	Starting wavelength.
5200	End wavelength.
0.03	Pixel size.
3.0e12	N(HI) lower cut-off for forest. No forest lines below this value included.
1.00e16	N(HI) upper limit for forest. No forest lines above this value included.
3.0	Spectral resolution, sigma, km/s.
100	Signal-to-noise ratio per pixel (but see above).
1	Parameter *inoise* in the code. See above.
2	Number of user-specified additional absorption systems (e.g. DLAs).
	You can specify 0 if you want. Max. coded is 20.
100	Avoidance zone around each additional system in km/s.
1.0e21	N(HI) of first system.
10.0	b-parameter of first system.
2.86618079	Redshift of first system.
1.0e21	N(HI) of second system.
10.0	b-parameter of second system.
2.86618079	Redshift of second system.
etc.	

References

[1] Penton, Stocke, Shull, ApJ Suppl., 152, 29, 2004