**Problem Definition (APPRIL project): One-Dimensional Seismic Monitoring**

**Stuart Russell, Computer Science Division, UC Berkeley**

**Simplified version: 7/8/14.**

The goal in this problem is to detect and localize seismic events given signals collected at detector stations over a fixed period.

In this initial version, the world is one-dimensional and the data are simulated. For a complete description of the real problem, see [Arora *et al*., (2013)](#http://www.cs.berkeley.edu/~russell/papers/bssa-netvisa.pdf). Subsequent versions will include spatial Gaussian process priors for seismicity, velocity, and absorptivity; a 2-D simulated world; and the 3-D real world.

The 1-D world’s spatial extent is the unit interval [0,1]. There are five detector stations *s*1,…, *s*5 with known locations *L*(*s*) = 0, 0.25, 0.5, 0.75, 1.0.

The generative model has two conceptual parts: the first part is a prior distribution over parameters describing the “physics” of the world, which are fixed but unknown; the second part describes events occurring in a particular episode (a time period of unit length) and any consequent detections. The prior over the physical parameters is as follows:

1. *λ*0, the overall rate of event occurrence, is drawn from Gamma(*αI*,*βI*).
2. Each station has a fixed, unknown background noise level *N*(*s*) drawn from an inverse gamma distribution InvGamma(*αN*,*βN*) and a fixed, unknown false alarm rate *F*(*s*) drawn from Gamma(*αF*,*βF*).
3. The signal velocity is a fixed unknown constant *V*0 = *W*02 where *W*0 is drawn from Normal(*μV*,*σ*2*V*). The travel time between two points is *S*(*x*,*y*) = |*x*-*y*|/*V*0.
4. The absorptivity per unit distance is a fixed unknown constant *α*0 = *β*02 where *β*0 is drawn from Normal(*μB*,*σ*2*B*).
5. The signal detection capability of each station is governed by fixed, unknown parameters *ν*(*s*),*σ* 2(*s*), with *ν*(*s*)drawn from Normal(*μv*,*σ* 2*v*) and *σ* 2(*s*) drawn from an inverse gamma InvGamma(*αs*,*βs*).
6. The arrival time measurement error at station *s* is governed by fixed, unknown parameters *μt*(*s*),*σ* 2*t*(*s*) drawn from a normal-inverse-gamma distribution with parameters (*μt*, λ*t*, *αt*, *βt*).
7. The amplitude measurement error at station *s* is governed by fixed, unknown parameters *μa*(*s*),*σ* 2*a*(*s*) drawn from a normal-inverse-gamma distribution with parameters (*μa*, λ*a*, *αa*, *βa*).
8. The amplitude distribution for noise detections at station *s* is governed by fixed, unknown parameters *μn*(*s*),*σ* 2*n*(*s*) drawn from a normal-inverse-gamma distribution with parameters (μ*n*, λ*n*, α*n*, β*n*).

For each episode the generative model is as follows:

1. The number of events occurring in the time period [0,1] is drawn from a Poisson with mean *λ*0. The location *L*(*e*) and the time *T*(*e*) of each event are drawn uniformly from [0,1].
2. The magnitude *M*(*e*) of each event is drawn from an exponential distribution with rate log 10; i.e., the probability density of magnitude *m* is proportional to 10-*m*. The minimum magnitude is 2.0.
3. The arriving log amplitude *A*(*e*,*s*) from event *e* at station *s* is given by *M*(*e*) - *α*0 |*L*(*e*) - *L*(*s*)| .
4. The probability that a signal from event *e* is detected at a station *s* is given by Logistic(*A*(*e*,*s*) - *N*(*s*),*ν*(*s*),*σ*(*s*)) and where Logistic(*x*,*ν*,*σ*) = 1/(1+exp(-(*x*-*ν*)/*σ*)).
5. If event *e* generates a detection *d* at station *s*, the detection attributes are as follows:
   1. The measured arrival time *t*(*d*) = *T*(*e*) + *S*(*L*(*e*), *L*(*s*)) + *εt* where *εt* ~ Normal(*μt*(*s*),*σ* 2*t*(*s*)).
   2. The measured log amplitude *a*(*d*) = *A*(*e*,*s*) + *εa* where *εa* ~ Normal(*μa*(*s*),*σ* 2*a*(*s*)).
   3. The sign *s*(*d*) indicating the direction of arrival of the signal is +1 if *L*(*e*) < *L*(*s*), -1 otherwise.
6. The number of noise detections *d* at station *s* is Poisson(*N*(*s*)):
   1. The measured arrival time *t*(*d*) is uniform in [0,1].
   2. The measured log amplitude *a*(*d*) ~ N(*μn*(*s*),*σ* 2*n*(*s*)).
   3. The sign *s*(*d*) indicating the direction of arrival of the signal is drawn uniformly from {+1,-1}.

For any given test, a subset of episodes (the training set) is available fully labeled, with the true events and associations between events and detections.In the test set, only the detections are available. The desired output for each period is a *bulletin*, i.e., a set of events with times, locations, and magnitudes.

To evaluate a bulletin, a matching is constructed between the bulletin and the ground truth for that period. A bipartite graph is created between predicted and true events. An edge is added between a predicted and a true event that is at most *d*max in distance and *t*max in time apart. The weight of the edge is the distance between the two events. Finally, a minimum weight–maximum cardinality matching is computed on the graph. Three metrics are extracted from this matching:

* Precision (percentage of predicted events that are matched);
* Recall (percentage of true events that are matched), and
* Average error (average distance between matched events).

The fixed, known parameters of the model are marked in red above. The values used for the current data set are as follows:

*αI* = 20 *βI* = 2

*αN* = 3 *βN* = 2

*αF* = 20 *βF* = 1

*μV* = 5 *σ*2*V* = 1

*μB* = 2 *σ*2*B* = 1

*μv* = 1 *σ*2*v* = 1

*αs* = 2 *βs* = 1

*μt* = 0 λ*t* = 1000 *αt* = 20 *βt* = 1

*μa* = 0 λ*a* = 1 *αa* = 2 *βa* = 1

*μn* = 0 λ*n* = 1 *αn* = 2 *βn* = 1