5512, Spring-2019
ASSIGNMENT 3:
Assigned: 03/11/19 Due: 03/31/19 at 11:55 PM (submit via Canvas, you may scan or take a picture of your paper answers) Submit only pdf or txt files (for non-code part), separate submission for code files
Show as much work as possible for all problems!

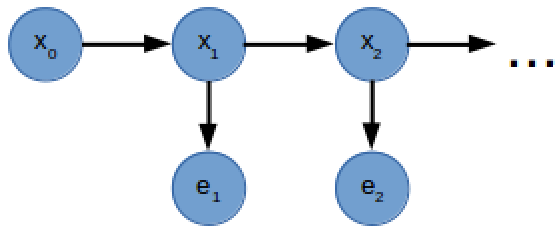
Problems 1 to 4 will use the following Bayesian network (specifically, a Hidden Markov Model): (Here X's represent whether the water table is low/medium/high and E's represent whether or not your basement gets flooded.)

$P(x_0)$

low	1/3
med	1/3
high	1/3

$P(x_{t+1}|x_t)$

from x_t	to x_{t+1}: low	med	high
low	0.6	0.35	0.05
med	0.2	0.6	0.2
high	0	0.5	0.5



$P(e_t|x_t)$

x_t	e_t seen: flooded	not flooded
low	0	1
med	0.05	0.95
high	0.4	0.6

Problem 1. (15 points)
Give the filtering probabilities for the water table's three values in the following cases:
(1) $P(x_1 \mid e_1=$"not flooded")
(2) $P(x_2 \mid e_1=$"not flooded", $e_2=$"not flooded")
(3) $P(x_3 \mid e_1=$"not flooded", $e_2=$"not flooded", $e_3=$"flooded")

Solution:
Here we use the forward message equation:
f_t+1 =α * p(e_t|x_t+1) sum_{x_t} P(x_t+1|x_t) f_t
... where f_0 = [1/3, 1/3, 1/3]

(1)
p(x_1=low | ~e_1) = 1 * (0.6 * 1/3 + 0.2*1/3 + 0*1/3) = 0.2667
p(x_1=med | ~e_1) = 0.95 * (0.35 * 1/3 + 0.6*1/3 + 0.5*1/3) = 0.4592
p(x_1=high | ~e_1) = 0.6 * (0.05 * 1/3 + 0.2*1/3 + 0.5*1/3) = 0.15

Normalize:
p(x_1=low | ~e_1) = 0.2667/(0.2667+0.4592+0.15) = 0.3045
p(x_1=med | ~e_1) = 0.4592/(0.2667+0.4592+0.15) = 0.5243
p(x_1=high | ~e_1) = 0.15/(0.2667+0.4592+0.15) = 0.1712

(2)
p(x_2=low | ~e_1, ~e_2) = 1 * (0.6 * 0.3045 + 0.2*0.5243 + 0*0.1712) = 0.2876
p(x_2=med | ~e_1, ~e_2) = 0.95 * (0.35 * 0.3045 + 0.6*0.5243 + 0.5*0.1712) = 0.4814
p(x_2=high | ~e_1, ~e_2) = 0.6 * (0.05 * 0.3045 + 0.2*0.5243 + 0.5*0.1712) = 0.1234

Normalize:
p(x_2=low | ~e_1, ~e_2) = 0.2876/(0.2876+0.4814+0.1234) = 0.3223
p(x_2=med | ~e_1, ~e_2) = 0.4814/(0.2876+0.4814+0.1234) = 0.5394
p(x_2=high | ~e_1, ~e_2) = 0.1234/(0.2876+0.4814+0.1234) = 0.1383

(3)
p(x_3=low | ~e_1, ~e_2, e_3) = 0 * (0.6 * 0.3223 + 0.2*0.5394 + 0*0.1383) = 0
p(x_3=med | ~e_1, ~e_2, e_3) = 0.05 * (0.35 * 0.3223 + 0.6*0.5394 + 0.5*0.1383) = 0.02528
p(x_3=high | ~e_1, ~e_2, e_3) = 0.4 * (0.05 * 0.3223 + 0.2*0.5394 + 0.5*0.1383) = 0.07726

Normalize:
p(x_3=low | ~e_1, ~e_2, e_3) = 0/(0+0.02528+0.07726) = 0
p(x_3=med | ~e_1, ~e_2, e_3) = 0.02528/(0+0.02528+0.07726) = 0.2465
p(x_3=high | ~e_1, ~e_2, e_3) = 0.07726/(0+0.02528+0.07726) = 0.7535

Problem 2. (15 points)
Given the same sequence of evidence as problem 1, find the smoothed estimates for x_1, x_2 and x_3. (In other words, the evidence is still: e_1=“not flooded”, e_2=“not flooded”, e_3=“flooded”)

Then plot the probabilities both filtering and smoothed estimates on the same graph. (Note: you will need two points/lines for a single probability, so overall you should have four lines.)

Solution:
We compute the backwards message as:
h_t-1 = sum_{x_t} P(e_t|x_t) P(x_t | x_{t-1}) h_t
... with h_3 = [1,1,1]

h_2(low) = (0*0.6*1 + 0.05*0.35*1 + 0.4*0.05*1) = 0.0375
h_2(med) = (0*0.2*1 + 0.05*0.6*1 + 0.4*0.2*1) = 0.11
h_2(high) = (0*0.0*1 + 0.05*0.5*1 + 0.4*0.5*1) = 0.225

h_1(low) = (1*0.6*0.0375 + 0.95*0.35*0.11 + 0.6*0.05*0.225) = 0.06582
h_1(med) = (1*0.2*0.0375 + 0.95*0.6*0.11 + 0.6*0.2*0.225) = 0.0972
h_1(high) = (1*0.0*0.0375 + 0.95*0.5*0.11 + 0.6*0.5*0.225) = 0.1198

... Now we multiply the backwards message with the corresponding filtering message (i.e. h_1 with P(x_1|~e_1)).

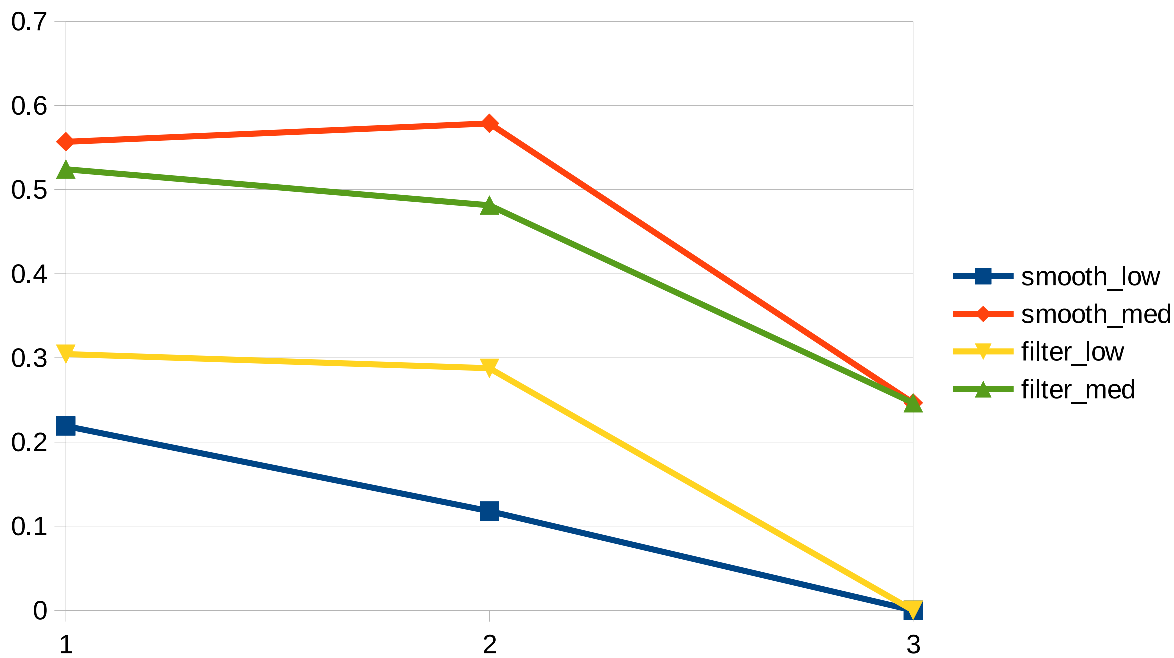
p(x_1=low | ~e_1, ~e_2, e_3) = 0.3045 * 0.06582 --> (normalize) = 0.3045 * 0.06582/(0.3045 * 0.06582+ 0.5243 * 0.0972 + 0.1712 * 0.1198) = 0.2190
p(x_1=med | ~e_1, ~e_2, e_3) = 0.5243 * 0.0972 --> (normalize) = 0.5243 * 0.0972/(0.3045 * 0.06582+ 0.5243 * 0.0972 + 0.1712 * 0.1198) = 0.5569
p(x_1=high | ~e_1, ~e_2, e_3) = 0.1712 * 0.1198 --> (normalize) = 0.1712 * 0.1198/(0.3045 *

0.06582+ 0.5243 * 0.0972 + 0.1712 * 0.1198) = 0.2241

p(x_2=low | ~e_1, ~e_2, e_3) = 0.3223 * 0.0375 --> (normalize) = 0.3223 * 0.0375/(0.3223 * 0.0375+ 0.5394 * 0.11+ 0.1383 * 0.225) = 0.1179
p(x_2=med | ~e_1, ~e_2, e_3) = 0.5394 * 0.11 --> (normalize) = 0.5394 * 0.11/(0.3223 * 0.0375+ 0.5394 * 0.11+ 0.1383 * 0.225) = 0.5787
p(x_2=high | ~e_1, ~e_2, e_3) = 0.1383 * 0.225 --> (normalize) = 0.1383 * 0.225/(0.3223 * 0.0375+ 0.5394 * 0.11+ 0.1383 * 0.225) = 0.3035

The smoothing estimate for x_3 is just the filtering estimate:
p(x_3=low | ~e_1, ~e_2, e_3) = 0
p(x_3=med | ~e_1, ~e_2, e_3) = 0.2465
p(x_3=high | ~e_1, ~e_2, e_3) = 0.7535

Plotting these probabilities (low+med):



Problem 3. (20 points)
Assume we are using the same HMM as problems 1 & 2, but we have difference evidence:
e_1=“flooded”, e_2=“not flooded”, e_3=“not flooded”, e_4=“flooded”, e_5=“not flooded”

What is the most likely sequence of water table levels for these five days?

If you found that day 6 was “flooded” (i.e. e_6=“flooded”), what is the most likely sequence now?

Solution:
We use the equation:
MLE(t) = p(e_t|x_t) * max(p(x_t|x_t-1) * MLE(t-1)), where MLE(0) = 1/3 = p(x_0)

To compute most likely sequence/explanation for day 1, we would find as:
x_1=low: 0 * max(0.6*1/3 + 0.2*1/3 + 0*1/3) = 0 (also record came from “low”, though it is a tie so

any is fine)
x_1=med: 0.05 * max(0.35*1/3 + 0.6*1/3 + 0.5*1/3) = 0.05 * 0.6*1/3 = 0.01 (also record came from "med")
x_1=high: 0.4 * max(0.05*1/3 + 0.2*1/3 + 0.4*1/3) = 0.4 * 0.4 * 1/3 = 0.5333 (also record came from "high"

... continue this process to get:

	x0	x1	x2	x3	x4	x5	x6
low	0.3333333333	0	0.002	0.0050666667	0	8.664E-05	0
med	0.3333333333	0.01	0.0253333333	0.01444	0.0004332	**0.00054872**	1.64616E-05
high	0.3333333333	0.0533333333	0.0128	0.003072	0.0011552	0.000277248	**4.435968E-05**

... with "came from" tables as:

	x1	x2	x3	x4	x5	x6
low						
	0.2	0	0.0012	**0.00304**	0	5.1984E-05
	0.0666666667	**0.002**	**0.0050666667**	0.002888	**8.664E-05**	**0.000109744**
	0	0	0	0	0	0
med						
	0.1166666667	0	0.0007	0.0017733333	0	3.0324E-05
	0.2	0.006	**0.0152**	**0.008664**	0.00025992	**0.000329232**
	0.1666666667	**0.0266666667**	0.0064	0.001536	**0.0005776**	0.000138624
high						
	0.0166666667	0	0.0001	0.0002533333	0	4.332E-06
	0.0666666667	0.002	0.0050666667	**0.002888**	8.664E-05	0.000109744
	0.1333333333	**0.0213333333**	**0.00512**	0.0012288	**0.00046208**	**0.0001108992**

... Since the highest number on day 5 is "med", we backtrack to see the med came from a "high" transition (bold number in middle rows for x5). Then in x4 we see "high" came from a "med" (bold number in bottom rows for x4)... and repeat this to get the sequence:

Day 5 sequence: reverse[med, high, med, med, high, high]
Day 5 sequence: high, high, med, med, high, med

Day 6 sequence: reverse [high, high, high, med, med, high, high]
Day 6 sequence: high, high, med, med, high, high, high

Problem 4. (25 points)
Use particle filtering to estimate:
P(x_{10} | e_1="not flooded", e_2="not flooded", e_3="flooded", e_4="flooded", e_5="not flooded", e_6="not flooded", e_7="not flooded", e_8="not flooded", e_9="flooded", e_{10}="not flooded")

(i.e. the days "flooded" are 3,4 and 9. The rest are "not flooded".) Give the number of particles used in your sampling, along with the probability for the water table values.

Solution:
With 1 million particles:
low: 0.061002
med: 0.620487
high: 0.318511

Problem 5. (20 points)
Assume we are using the Frisbee example from class, where: $P(x_0) = N(0,1)$ and $P(e_t|x_t) = N(x_t, 0.75)$. How accurate do you need to be so that after 10 throws, the variance is not more than 10.

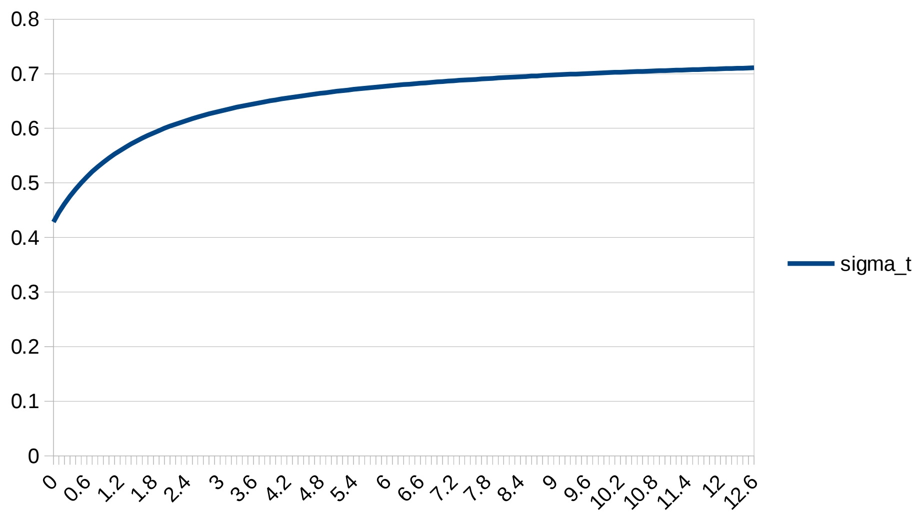
Solution:
This is actually not possible. You can reason this, that even if the throw was completely inaccurate (i.e. P(x_t+1 | x_t) = N(0, ∞), then the evidence (i.e. P(e_t | x_t)), would still bound it by N(x_t, 0.75). You can either directly show this or approximate it using the equation:

$$\sigma^2_{t+1} = \frac{(\sigma^2_t + \sigma^2_x)\cdot\sigma^2_z}{\sigma^2_t + \sigma^2_x + \sigma^2_z}$$

So for our numbers we get:

$$\sigma^2_{t=1} = \frac{(1+\sigma^2_x)\cdot 0.75}{1+\sigma^2_x+0.75}$$

You can graph this as:

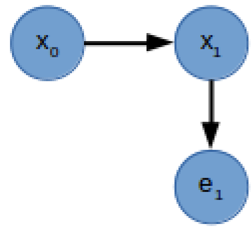


Or do it mathematically (L'Hôpital's rule):

$$\sigma^2_{t=1} = \lim_{\sigma^2_x \to \infty} \frac{(1+\sigma^2_x)\cdot 0.75}{1+\sigma^2_x+0.75} = \frac{0.75\cdot\sigma^2_x}{\sigma^2_x} = 0.75$$

Since on day1 the variance is capped at 0.75, by day 10, the most variance you can have is 0.500000000573594

Problem 6. (15 points)
Suppose you have the simple HMM 3-node Bayes net shown below. Assume that $P(x_0)$ is uniformly distributed between [-1, 1]. The transition probability, $P(x_{t+1} | x_t)$, is $N(x_t,1)$. The evidence probability is: $P(e_t|x_t) = N(x_t, 0.5)$



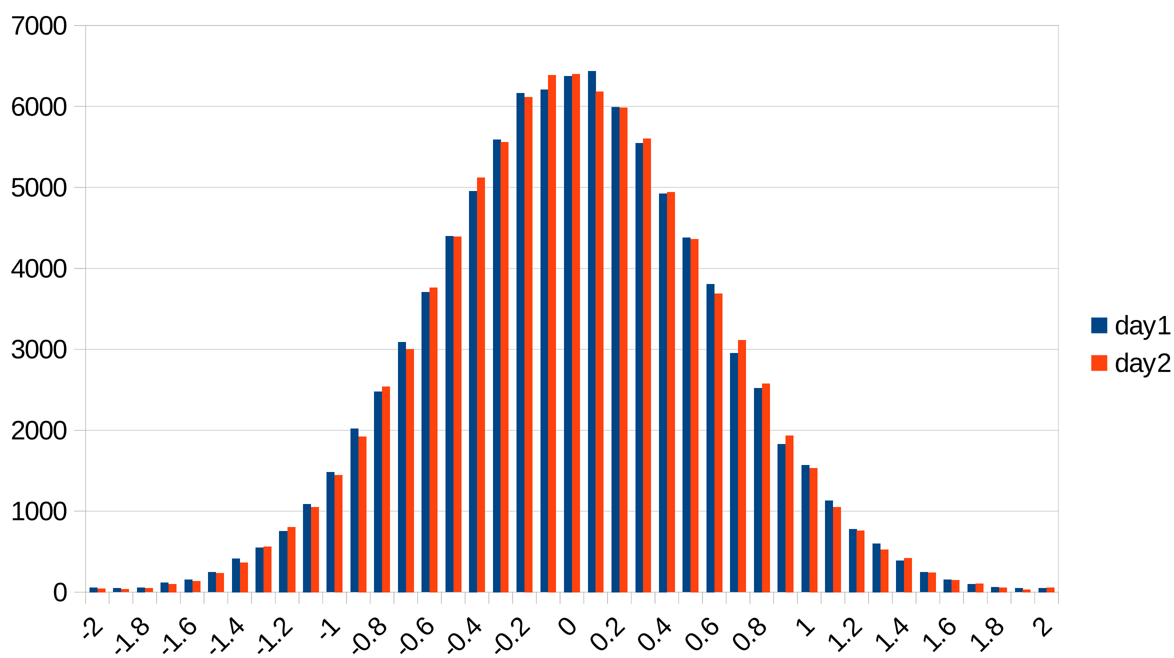
Use this description to approximate the distribution for $P(x_1 \mid e_1=0)$ and plot this distribution on a graph. Extend this network to 5 nodes to estimate the distribution for $P(x_2 \mid e_1=0, e_2=0)$ and plot this distribution on a graph. As $t \to \infty$, describe what happens to the distribution of the filtering probabilities (assuming the evidence always says zero).

Solution:
You could setup this as an integral to try and find the exact answer, but this is quite difficult. Instead, I would treat this as a sort of "particle filtering" where you just sample to estimate the distribution (since this is continuous, we will discretize it like a histogram).

Process:
1. Generate a random sample from uniform [0,1]
2. Apply the N(sample, 1) to move the particle somewhere (again sampling)
3. Weight this sample
4. Do a lot, and then re-sample (this is day-1 distribution)

Then we repeat to get day-2 distribution, except step 1 is from our distribution generated for day-1

This gives the following distribution for days 1 and 2 (already pretty normal):



As $t \to \infty$, the initial distribution will not have an effect, and the transition + evidence normals will dominate the distribution. Specifically, it will converge for when:

$$\sigma_t^2 = \frac{(\sigma_t^2+\sigma_x^2)\cdot\sigma_z^2}{\sigma_t^2+\sigma_x^2+\sigma_z^2} = \frac{(\sigma_t^2+1)\cdot 0.5}{\sigma_t^2+1+0.5}$$

Solving for $\sigma_t^2 = 0.366025403784439$.

For the mean we have:

$$\mu_{t+1} = \frac{(\sigma_t^2+\sigma_x^2)\cdot z_{t+1}+\sigma_x^2\cdot\mu_t}{\sigma_t^2+\sigma_x^2+\sigma_z^2} = \frac{(0.366+1)\cdot 0+1\cdot\mu_t}{0.366+1+.5}$$

The numerator is just the old mean, and the denominator is greater than zero. So as $t \to \infty$, the mean should go towards zero. So it should stabilize to N(0, 0.366).