1-D Tracking

Imagine you have a self driving car with one LiDAR sensor. A LiDAR sensor is a way to measure distance to other objects. You don't need to know how LiDAR works, just that it will give you a measure of distance. You are trying to detect how far away an object is from the self driving car.

Let T be the true distance from the car to the object. Our prior belief is that $T \sim N(\mu = 1, \sigma^2 = 3)$.

Our LiDAR sensor can measure distance but it isn't perfectly accurate. It has some noise due to measurement error within the instrument. Let X be the distance given by the LiDAR. We say that X is equal to the true distance plus some noise: $X = t + \text{Noise}$. Let M be the noise and $M \sim N(\mu = 0, \sigma^2 = 1.5)$.

a. What is the likelihood function $f(X = x|T = t)$? This is asking, if you knew a value for t, how could you express X. Specifically, use linearity of expectation and linearity of variance to find the parameters of this distribution.

b. We observe a LiDAR measurement of 4 meters. Write out the equation for the probability density $f(X = 4|T = t)$.

c. We want to update our belief in the true distance to the object $f(T = t|X = 4)$. Write your answer in terms of a constant K that you do not need to solve for.

CS109 Lecture 12 Worksheet: General Inference

Bayesian Networks, Generative Models, and Sampling

Summer 2026

Lecture 12 Worksheet: General Inference

Topics: Bayesian networks (probabilistic graphical models), the joint as a product of conditional probabilities, conditional independence of a node from its non-descendants given its parents, generative/ancestral sampling from the joint, and approximate inference by rejection sampling (with a note on why we eventually need MCMC).

How to use this sheet. Work in small groups. A TA and the instructor will circulate to help. We will go over selected solutions on the board. You do not need to finish every part; aim for understanding over speed. Keep fractions exact where you can.

Problem 1 (Review of Lecture 11): One More Update

A spam filter has prior $P(\text{Spam}) = 0.3$. A message contains the word "free," where $P(\text{"free"} \mid \text{Spam}) = 0.6$ and $P(\text{"free"} \mid \text{Not Spam}) = 0.1$.

(a) Compute the posterior $P(\text{Spam} \mid \text{"free"})$.

The Network for Problems 2–5

Consider a small "WebMD" Bayesian network with three binary random variables and edges Flu $\to$ Fever and Flu $\to$ Tired:

- $P(\text{Flu} = 1) = 0.2$
- $P(\text{Fever} = 1 \mid \text{Flu} = 1) = 0.9$, $\quad P(\text{Fever} = 1 \mid \text{Flu} = 0) = 0.05$
- $P(\text{Tired} = 1 \mid \text{Flu} = 1) = 0.8$, $\quad P(\text{Tired} = 1 \mid \text{Flu} = 0) = 0.3$

Problem 2: The Joint as a Product

(a) Using the network structure, write the joint $P(\text{Flu}, \text{Fever}, \text{Tired})$ as a product of conditional probabilities.

(b) Compute $P(\text{Flu} = 1, \text{Fever} = 1, \text{Tired} = 0)$.

(c) A full joint table over 3 binary variables has how many rows? Over n binary variables? Why does the network's factorization save space in general?

Problem 3: Conditional Independence

(a) The network encodes that a node is conditionally independent of its non-descendants given its parents. Which conditional independence does this network assert about Fever and Tired?

(b) Compute $P(\text{Fever} = 0 \mid \text{Flu} = 1)$.

(c) Is $P(\text{Fever} = 1 \mid \text{Tired} = 1)$ equal to $P(\text{Fever} = 1)$? Answer conceptually: are Fever and Tired *marginally* independent, even though they are conditionally independent given Flu?

Problem 4: Sampling from the Joint (Generative Model)

(a) Describe an algorithm (ancestral sampling) to draw one sample $(\text{Flu}, \text{Fever}, \text{Tired})$ from the joint, sampling variables in an order consistent with the arrows.

(b) Why is "being able to write a program that samples from the joint" the same as "having defined the joint"?

Problem 5: Rejection Sampling

You want $P(\text{Flu} = 1 \mid \text{Fever} = 1)$ by rejection sampling.

(a) Describe the rejection-sampling algorithm: which samples do you keep, and how do you form the estimate?

(b) Explain why this estimate approximates the true conditional probability. (Hint: probability as a long-run frequency.)

(c) Compute the **exact** $P(\text{Flu} = 1 \mid \text{Fever} = 1)$ with Bayes' theorem so you know what rejection sampling should converge to.

Problem 6: Rejection Sampling on a Bigger Network (pset4: web_md)

A 10-variable "WebMD" network models Lyme disease and Flu from risk factors (tick bite, stress) and symptoms (fever, rash, tired, cough, runny nose, headache). Each variable is Bernoulli with a probability determined by its parents. You implement rejection sampling to estimate, for the observation $\{\text{fever} = 1, \text{tick} = 1, \text{cough} = 0\}$:

$$P(\text{Lyme} = 1 \mid \text{fever} = 1, \text{tick} = 1, \text{cough} = 0) \quad \text{and} \quad P(\text{Flu} = 1 \mid \cdots).$$

(a) Outline how one joint sample is generated from this network (which variables are sampled first, and how each variable's probability is chosen).

(b) Describe the reject/keep rule for this observation and how the two target probabilities are estimated. Why might they **not** sum to 1?

Challenge Problem (optional): When Rejection Sampling Breaks

(a) Suppose the observed evidence is very unlikely (say it occurs in 1 of every 10,000 joint samples). What happens to the efficiency of rejection sampling, and why?

(b) In one or two sentences, explain how MCMC (e.g. sampling with the observed variables held fixed) addresses this problem, at a conceptual level.