Basic Probability

1

Consider an organism with a genome in which mutations happen as a Poisson process with rate ν. Assume the following about mutations:

- All mutations are neutral (i.e., they do not affect the rate of reproduction).
- The genome is large enough that the mutations always happen at different loci (this is known as the infinite sites model) and are irreversible.
- At $t = 0$, there are no mutations.

a) What is the probability that the genome does not obtain any new mutations within the time interval $[0, t)$?

b) What is the expected number of mutations for time T?

c) Consider a population of N individuals following the Wright-Fisher model. Each generation is formed from the previous by the following algorithm:

- Every individual organism reproduces asexually (i.e., every organism divides forming two new organisms). The population size is now $2N$.
- Uniformly at random sample N of these $2N$ individuals to survive. This is now the new generation.

Find the probability $P(t)$ of two individuals having their "first" (latest chronologically) common ancestor t generations ago. Hint: Go backwards in time with discrete time steps. What is the continuum limit of this probability (i.e., the result for a large population size)?

d) Now add mutations to the Wright-Fisher model. Assume we sample two individuals that followed two distinct lineages for precisely t generations (i.e., their first common ancestor occurred t generations ago). What is $P(\pi|t)$, the probability of π mutations arising during the t generations?

e) What is $P(\pi)$, the probability of two individuals being separated by π mutations after they were born from the same parent? What is the expected value of π? (You may work in the continuum limit as in (c), corresponding to a large population size $N \gg 1$).

a) $p(k) = \frac{v^k e^{-v}}{k!}$

$p(0) = \frac{v^0 e^{-v}}{0!} = e^{-v}$

b) $v = \frac{d\,\text{mutations}}{dt}$

vT = mutations

c) P that 2 individuals i & j have same parent

P(i & j siblings | i survives & selected) = P(sibling j survives) × P(sibling j selected)

$P(\text{sibling } j \text{ survives}) = \frac{N - \text{sibling } i}{2N - \text{sibling } i} = \frac{N-1}{2N-1}$

$P(\text{sibling } j \text{ selected}) = \frac{1}{N-1}$

$P(i \,\&\, j \text{ siblings} \mid i \text{ survives} \,\&\, \text{selected}) = \frac{N-1}{2N-1} \frac{1}{N-1} = \frac{1}{2N-1}$

Therefore $P(\text{individuals } i \,\&\, j \text{ are not siblings}) = 1 - \frac{1}{2N-1} = \frac{2N-2}{2N-1}$

P(i & j having their first ancestor t generations ago)

$= P(i\,j \text{ not siblings @ each gen after gen} = t \text{ gens ago})^{t-1} \times P(i\,j \text{ siblings @ gen} = t \text{ gens ago})^1$

$= \left(\frac{2N-2}{2N-1}\right)^{t-1} \left(\frac{1}{2N-1}\right)$

for $N >> 1$

$(1)^{t-1} \frac{1}{2N-1} = \frac{1}{2N-1}$

d) $P(i\text{'s mutations} + j\text{'s mutations} = \pi)$

$= \sum_a^{\pi} p(a)\, p(\pi - a)$

$= \sum_a^{\pi} \frac{(vt)^a e^{-vt}}{a!} \frac{(vt)^{\pi-a} e^{-vt}}{(\pi - a)!}$

$= \sum_a^{\pi} \frac{(vt)^{\pi} e^{-2vt}}{a!\,(\pi - a)!}$

$= (vt)^{\pi} e^{-2vt} \frac{2^{\pi}}{\pi!}$

e) same parents = they are siblings, so only 1 gen away. $t=1$

$P(\pi) = (vt)^{\pi} e^{-2vt} \frac{2^{\pi}}{\pi!}$

$= v^{\pi} e^{-2v} \frac{2^{\pi}}{\pi!} = e^{-2v} \frac{(2v)^{\pi}}{\pi!} = \text{Pois}(\lambda = 2v,\, x = \pi)$

expected value:

$E[\pi] \sim \text{Pois}(\lambda = 2v) = 2v$

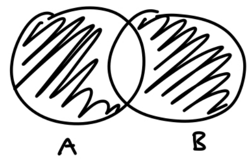
2

Bonferroni's inequality.

Prove that for any two events A and B, we have

$$P(A \cap B) \geq P(A) + P(B) - 1$$

$1 \geq P(A) + P(B) - P(A \cap B)$



$= P(A \cup B) \leq 1$ ✓

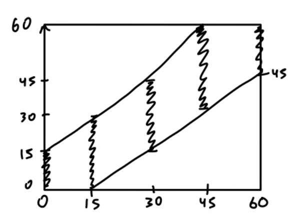
3

Romeo and Juliet have a date.

Each will arrive at the meeting place with a delay between 0 and 1 hour, with all pairs of delays being equally likely.

The first to arrive will wait for 15 minutes and will leave if the other has not yet arrived.

Question: What is the probability that they will meet?



normalized to 1

$1^{st}\ \Delta = \frac{1}{2} 0.75 \times 0.75$

$2^{nd}\ \Delta = \frac{1}{2} 0.75 \times 0.75$

$P(\text{meet}) = 1 - 0.75 \times 0.75$

4

A parking lot contains 100 cars, k of which happen to be lemons. We select m of these cars at random and take them for a test drive. Find the probability that n of the cars tested turn out to be lemons.

$$\frac{\binom{k}{n}\binom{100-k}{m-n}}{\binom{100}{m}}$$

5

Consider ten independent rolls of a 6-sided die. Let X be the number of $6s$ and let Y be the number of $1s$ obtained.

What is the joint PMF of X and Y?

$$= P(X)\,P(Y|X)$$

possible extracted sequenus of events

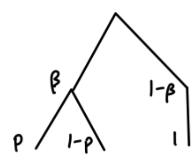
$$P(X) = \binom{10}{X} \underbrace{\left(\frac{1}{6}\right)^X \left(\frac{5}{6}\right)^{10-X}}_{\text{P of 1 extracted sequence of events}}$$

$$P(Y|X) = \binom{10-X}{Y} \left(\frac{1}{5}\right)^Y \left(\frac{4}{5}\right)^{10-X-Y}$$

$$P(X)\,P(Y|X) = \binom{10}{X}\left(\frac{1}{6}\right)^X\left(\frac{5}{6}\right)^{10-X} \binom{10-X}{Y}\left(\frac{1}{5}\right)^Y\left(\frac{4}{5}\right)^{10-X-Y}$$

6

We know that a treasure is located in one of two places, with probabilities β and $1-\beta$, respectively, where $0 < \beta < 1$. We search the first place and if the treasure is there, we find it with probability $p > 0$.
Show that the event of not finding the treasure in the first place suggests that the treasure is in the second place

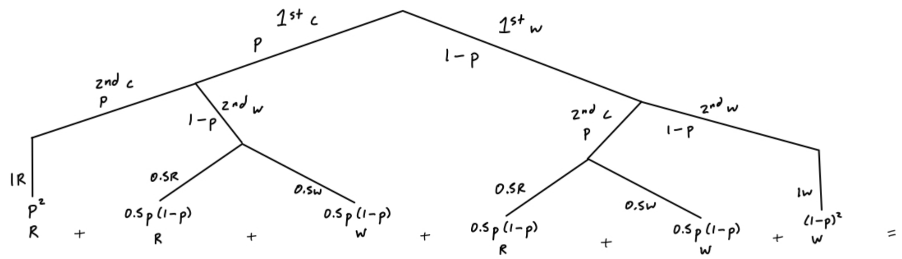


$$p\beta + (1-p)\beta + (1-\beta) = 1$$

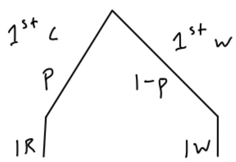
$$P(@\ 2^{nd} \mid \text{not found}) = \frac{P(@\ 2^{nd}\ \&\ \text{not found})}{P(\text{not found})} = \frac{1-\beta}{(1-p)\beta + (1-\beta)} > P(@\ 2^{nd} \mid \text{found or not found}) = 1-\beta$$

7

A hunter has two hunting dogs. On a two-path road, each dog will choose the correct path with probability p independently. Let each dog choose a path, and if they agree, take that one, and if they disagree, randomly pick a path. Is this strategy better than just letting one of the two dogs decide on a path?



$P(R) = p^2 + 0.5\,p(1-p) + 0.5\,p(1-p) = p^2 + p - p^2 = p$



$P(R) = p$

8

There are n different power plants. And each one can produce enough electricity to supply the entire city. Now suppose that the ith power plants fails independently with probability p_i.

(a) What is the probability that the city will experience a black-out?

a) $\prod_{n}^{N} p_n$

b) $\prod_{n}^{N} p_n + \sum_{s}^{N} \left[(1-p_s) \prod_{n \neq s}^{N-1} p_n \right]$

Master Equations and Chapman-Kolmogorov

9

The SIR model describes the evolution of four populations (compartments) over time: susceptible individuals (S) can be infected in contact with infectious individuals (I) with rate β/N. Once infected, the susceptible individuals become become infectious (I). Furthermore, infected individuals become removed (R) at a rate γ. Assuming a well mixed population, the SIR compartment model consists of the following reactions:

$$S + I \xrightarrow{\beta/N} I,$$
$$I \xrightarrow{\gamma} R.$$

b) Write down the master equation for the above process. Use the master equation to derive the ordinary differential equation (ODE) for the expected compartment population sizes.

$B = \beta/N$

b) $$\frac{\partial P(s,i,r)}{\partial t} = B(s+1)i\,P(s+1,i,r) - B\,s\,i\,P(s,i,r) + \gamma(i+1)P(s,i+1,r-1) - \gamma\,i\,P(s,i,r)$$

$$\frac{\partial S}{\partial t} = -B\,S\,I$$

$$\frac{\partial I}{\partial t} = -\gamma I$$

$$\frac{\partial R}{\partial t} = \gamma I$$

10

The asymmetric random walk is governed by the master equation

$$\frac{dp_n}{dt} = \alpha p_{n+1} + \beta p_{n-1} - (\alpha+\beta)p_n,$$

where n is the position. Find $\langle n\rangle$ and $\langle n^2\rangle$.

$$\frac{d\langle n\rangle}{dt} = \sum_n \Big[(n-1)\alpha\, p(n) + (n+1)\beta\, p(n) - n(\alpha+\beta)\, p(n)\Big]$$

$$= -\alpha + \beta$$

$$\langle n\rangle = (-\alpha+\beta)t$$

$$\frac{d\langle n^2\rangle}{dt} = \sum_n \Big[(n-1)^2\alpha\, p(n) + (n+1)^2\beta\, p(n) - n^2(\alpha+\beta)\, p(n)\Big]$$

$$= \sum_n \Big[(n^2-2n+1)\alpha\, p(n) + (n^2+2n+1)\beta\, p(n) - n^2(\alpha+\beta)\, p(n)\Big]$$

$$= \sum_n \Big[(-2n+1)\alpha\, p(n) + (2n+1)\beta\, p(n)\Big]$$

$$= 2\langle n\rangle(\beta-\alpha) + (\alpha+\beta)$$

$$= 2(\beta-\alpha)t + (\alpha+\beta)$$

$$\langle n\rangle = 2(\beta-\alpha)^2 t^2 + (\alpha+\beta)t$$

11

Consider the following reactions:

$$A \xrightarrow{\nu} B,$$
$$B \xrightarrow{\mu} A,$$

where the quantity above the arrows are the rates of each reaction. We denote a as the number of species A and b as the number of species B.

a) **[4 points]** Write down the master equation corresponding to this system.

b) **[8 points]** Derive the ODE for $\langle a\rangle$ and $\langle b\rangle$ from the master equation, where $\langle\cdot\rangle$ denotes the expected value.

c) **[8 points]** Noticing that $a+b=n$ is constant over time, express $\langle a\rangle$ and $\langle b\rangle$ at equilibrium in terms of n, ν and μ.

a) $$\frac{dP(a,b)}{dt} = v(a+1)\,P(a+1,b-1) - va\,P(a,b) + \mu(b+1)\,P(a-1,b+1) - \mu b\,P(a,b)$$

b) $$\frac{d\langle a\rangle}{dt} = \frac{d}{dt}\sum_a \sum_b \Big[(a-1)\,v a\,P(a,b) - v a^2 P(a,b) + (a+1)\,\mu b\,P(a,b) - a\,\mu b\,P(a,b) \Big]$$

$$= \frac{d}{dt}\sum_a \sum_b \Big[-va\,P(a,b) + \mu b\,P(a,b) \Big]$$

$$= -v\langle a\rangle + \mu\langle b\rangle$$

b is the same, but flipped

$$\frac{d\langle b\rangle}{dt} = -\mu\langle b\rangle + v\langle a\rangle$$

c) $b = n - a$

$$\frac{d\langle a\rangle}{dt} = -v\langle a\rangle + \mu n - \mu\langle a\rangle$$

$$\frac{d\langle a\rangle}{dt} + (v+\mu)\langle a\rangle + \mu n = 0$$

int factor $= e^{\int (v+\mu)\,dt} = e^{(v+\mu)t}$

$$e^{(v+\mu)t}\langle a\rangle = \int \mu n\, e^{(v+\mu)t}\,dt$$

$$\langle a\rangle = \mu n \frac{1}{(v+\mu)}$$

$$\langle b\rangle = n - \frac{\mu n}{v+\mu}$$

12

The following Python program implements the Metropolis-Hastings algorithm:

```
import random
f = [0, 0.1, 0.8, 0.1, 0]
x = 1
while True:
    y = x + random.choice([1, -1])
    if f[y] > f[x] * random.uniform(0, 1):
        x = y
        print(x)
```

It was designed to print numbers 1, 2, and 3 with frequencies 0.1, 0.8, and 0.1, respectively, but it has an error.

a) **[12 points]** Find the frequencies of the numbers printed by the incorrect program.

b) **[8 points]** Correct the error in the program.

a) doesn't print unless passes,

so printed #'s not based on prob listed

$\rightarrow$ all equal transitional probabilities

$$\boxed{1} \leftrightarrow \boxed{2} \leftrightarrow \boxed{3}$$

$p_1 + p_2 + p_3 = 1$

detailed balance

$$p_1 = p_3$$

two ways to get to state 2 in the network

2x as much time in state 2

$$p_1 = p_2/2$$

$p_1 = \frac{1}{4}$ $p_2 = \frac{1}{2}$ $p_3 = \frac{1}{4}$

b) print whether pass or not

13

Each day, the price of a stock either goes up, down, or stays constant. On Monday, Tuesday, and Wednesday, the stock price follows one of the patterns below with equal probability (other patterns are not possible, assume they have zero probability):

```
up, down, down
down, up, down
down, down, up
up, up, up
```

Consider a daily price move as a state of a random process.

a) Show that the process is not Markovian.

b) Show that the process obeys the Chapman-Kolmogorov equation.

a) $P(u|u) = \frac{2}{1+1+2} = 0.5$

$P(u|u,u) = 1.0$

$P(u|u) \neq P(u|u,u)$

not Markovian

b) $P(u|u) = 0.5 = 1 - \overset{0.5}{P(d|u)}$

$P(u|d) = \frac{1+1}{1+1+2} = 0.5 = 1 - \overset{0.5}{P(d|d)}$

$\overset{P(u,u,u)}{P(u|u)\,P(u|u)} + P(u|d)\,P(d|u) = 0.5 = \overset{P(u,u)}{P(u|u)}$ ✓

$P(u|u)\,P(u|d) + P(u|d)\,P(d|d) = 0.5 = P(u|d)$ ✓

all 1 step = 2 steps so Chapman-Kolmogorov

14

You toss a biased coin. The probability of head is $0 < q < 1$. Define a random process $\{X_n\}$ such that $X_n = 1$ if the $n-1$-th and the n-th toss are heads (two heads in a row), and $X_n = 0$ otherwise. Is this processes Markovian? Does it obey the Chapman-Kolmogorov equation?

a) $P(1|1) = q$ $\quad \underline{1}\ \underline{1}\ \underline{\ \ }$

$P(1|1,1) = q$ $\quad \underline{1}\ \underline{1}\ \underline{1}\ \underline{\ \ }$ ok

$P(1|0,1) = q$ $\quad \underline{0}\ \underline{1}\ \underline{1}\ \underline{\ \ }$ ok

$P(1|0) = \frac{1}{3} q$ $\quad \underline{0}\ \underline{1}\ \underline{\ \ }$

$\underline{0}\ \underline{0}\ \underline{\ \ }$

$\underline{1}\ \underline{0}\ \underline{\ \ }$

$P(1|1,0) = 0$ $\quad \underline{1}\ \underline{1}\ \underline{0}\ \underline{\ \ }$

$P(1|0,0) = \frac{2}{6} q$ $\quad \underline{0}\ \underline{1}\ \underline{\ \ }$ $<^{0\ \ 0q}_{1\ \ 0q}$

$\underline{0}\ \underline{0}\ \underline{\ \ }$ $<^{0\ \ 0q}_{1\ \ qq}$

$\underline{1}\ \underline{0}\ \underline{\ \ }$ $<^{0\ \ 0q}_{1\ \ qq}$

$P(1|0) \neq P(1|0,1)$

not Markovian

b) $P(1,1) = q$

$P(1,1)\,P(1,1) + P(1,0)\,P(0,1)$

$= q^2 + \frac{1}{3} q(1-q)$

$= \frac{2}{3} q^2 + \frac{1}{3} q \neq q$

not Chapman-Kolmogorov

15

Consider the predator-prey system

$$X \xrightarrow{\alpha} 2X,$$
$$X + Y \xrightarrow{\beta} 2Y,$$
$$Y \xrightarrow{\gamma} \emptyset,$$

where X and Y denote the population of rabbits and foxes, respectively.

Consider an initial population of $X(0) = 1000$ and $Y(0) = 1000$ with $\alpha = 10$, $\beta = 0.01$ and $\gamma = 10$.

a) Derive the ODE corresponding to the above system. What is the solution of the ODE?

$$\frac{dx}{dt} = \alpha x - \beta x y$$

$$\frac{dy}{dt} = \beta x y - \gamma y$$

$$\frac{dx}{dt}(0) = 10\,(1000) - 0.01\,(1000 \times 1000) = 0$$

$$\frac{dy}{dt}(0) = 0$$

no change. $x(t) = 1000$ $y(t) = 1000$

Bayesian Inference

16

You are in charge of determining the probability θ that produced parts of a selected machine pass a quality assessment test. For this purpose, you randomly select 2 parts from the daily production of this machine for ten days and examine them. All selected part (your data D) pass the quality test.

a) **[2 points]** Determine the Maximum-Likelihood-Estimate for θ given the data D.

b) **[7 points]** As a prior for the probability θ you are selecting the following distribution

$$f(\theta) = \begin{cases} C\,\theta^3\,\exp(-2\theta) & \text{for } \theta \in [0,1] \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

with the normalizing constant C. Determine the Maximum-A-Posteriori-Estimate for θ given the data D and the prior $f(\theta)$.

c) **[8 points]** You decide to use a parameterized prior $g(\theta|a)$ instead of $f(\theta)$:

$$g(\theta|a) = \begin{cases} (a+1)\,\theta^a & \text{for } \theta \in [0,1] \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

Determine the A-Posteriori-Density $f(\theta|D,a)$ given the data D and the prior $g(\theta|a)$.

d) **[3 points]** You are randomly selecting another part produced by the machine. Based on the A-Posteriori-Density found in the previous subquestion, what is the expected probability that the part passes the quality assessment test ?

a) $\theta = P(1 \text{ part passes})$ $\quad D = \{x_1 \dots x_{20}\}$

$X \sim \theta^x (1-\theta)^x$

$P(D|\theta) = \prod_n^{20} \theta^{x_n}(1-\theta)^{1-x_n}$

$\theta_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln P(D|\theta)$

$\ln P(D|\theta) = \sum_n^{20} x_n \ln\theta + \sum_n^{20}(1-x_n)\ln(1-\theta)$

$0 = \frac{\partial}{\partial\theta}\ln P(D|\theta) = \sum_n x_n \frac{1}{\theta} - \sum_n (1-x_n)\frac{1}{1-\theta}$

$\sum_n x_n - \theta\sum_n x_n = \sum_n (1-x_n)\theta = 20\theta - \sum_n x_n\theta$

$\theta_{MLE} = \frac{1}{20}\sum_n^{20} x_n$ makes sense.

all parts pass so all $x_n = 1$

$\boxed{\theta_{MLE} = 1}$

$\frac{\partial^2}{\partial\theta^2}\ln P(D|\theta) = \sum_n x_n \frac{-1}{\theta^2} - \sum_n (1-x_n)\frac{+1}{(1-\theta)^2}$ ✓ always neg approaching $\theta = 1$

b) $\theta_{MAP} = \underset{\theta}{\operatorname{argmax}}\, P(\theta|D) = \underset{\theta}{\operatorname{argmax}}\left[\ln P(D|\theta) + \ln P(\theta)\right]$

$\ln P(\theta|D) = \ln P(D|\theta) + \ln f(\theta)$

$0 = \sum_n x_n \frac{1}{\theta} - \sum_n (1-x_n)\frac{1}{1-\theta} + \frac{\partial}{\partial\theta}\left[\ln C + 3\ln\theta - 2\theta\right]$

$= \sum_n x_n \frac{1}{\theta} - \sum_n (1-x_n)\frac{1}{1-\theta} + \frac{3}{\theta} - 2$

$= \left(\sum_n x_n + 3 - 2\theta\right)(1-\theta) = \theta\sum_n(1-x_n)$

$\sum_n x_n + 3 - 2\theta - \theta\sum_n x_n - 3\theta + 2\theta^2 = 20\theta - \sum_n x_n\theta$

$\sum_n x_n + 3 - 25\theta + 2\theta^2 = 0$

all parts pass so all $x_n = 1$

$2\theta^2 - 25\theta + 23 = 0$

$(\theta - 1)(2\theta - 23) = 0$

$\boxed{\theta_{MAP} = 1}$

$\frac{\partial^2}{\partial\theta^2} = -1(23 - 2\theta) + (-2)(1-\theta) - \sum_n(1-x_n)$ ✓ neg

c) $P(\theta|D,a) = \frac{P(D|\theta,a)\,P(\theta|a)}{P(D|a)}$

$\propto \prod_n^{20}\theta^{x_n}(1-\theta)^{1-x_n}\quad(a+1)\theta^a$

$P(\theta|D,a) = \frac{\prod_n^{20}\theta^{x_n}(1-\theta)^{1-x_n}\,(a+1)\theta^a}{\int_0^1\prod_n^{20}\theta^{x_n}(1-\theta)^{1-x_n}\,(a+1)\theta^a\,d\theta}$

all parts pass so all $x_n = 1$

$= \frac{\theta^{20}(a+1)\theta^a}{\int_0^1\theta^{20}(a+1)\theta^a\,d\theta}$

$= \frac{(a+1)\,\theta^{20+a}}{(a+1)\int_0^1\theta^{20+a}\,d\theta}$

$= \frac{\theta^{20+a}}{\frac{1}{21+a}\left[\theta^{21+a}\right]_0^1}$

$= \boxed{(21+a)\,\theta^{20+a}}$

d) $\theta_{MAP} = \underset{\theta}{\operatorname{argmax}}\left[P(D|\theta)(21+a)\,\theta^{20+a}\right]$

$0 = \sum_n x_n\frac{1}{\theta} - \sum_n(1-x_n)\frac{1}{1-\theta} + \frac{\partial}{\partial\theta}\left[\ln(21+a) + (20+a)\ln\theta\right]$

$= \sum_n x_n\frac{1}{\theta} - \sum_n(1-x_n)\frac{1}{1-\theta} + (20+a)\frac{1}{\theta}$

$\frac{\sum_n x_n + 20 + a}{\theta} = \frac{\sum_n(1-x_n)}{1-\theta}$

$\sum_n x_n + 20 + a - \cancel{\sum_n x_n\theta} - 20\theta - a\theta = 20\theta - \cancel{\sum_n x_n\theta}$

$\sum_n x_n + 20 + a = (40+a)\theta$

all parts pass so all $x_n = 1$

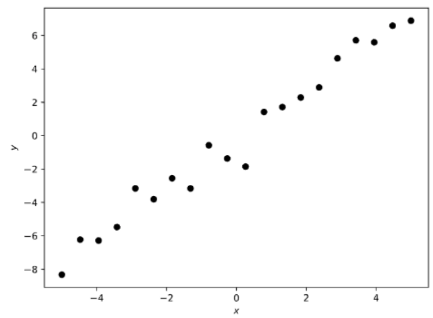
$\theta_{MAP} = \frac{40+a}{(40+a)} = 1$

$\boxed{100\% = \text{expected probability of passing}}$

$\frac{\partial^2}{\partial\theta^2} = \sum_n x_n\frac{-2}{\theta^2} - \sum_n(1-x_n)\frac{2}{\theta^2} - (20+a)\frac{2}{\theta^2}$ ✓

17

An experimentalist collected data $D = \{x_i, y_i\}_{i=1}^{N}$, reported on the figure below (also available in the file `dataQ2.csv`).



A model of this data is given by

$$y_i = ax_i + \sigma\xi_i,$$

where a and σ are scalars and ξ_i are i.i.d. normally distributed random variables, $\xi_i \sim \mathcal{N}(0, 1)$.

a) **[10 points]** Write the log-likelihood of the data given the parameters $\theta = (a, \sigma)$. Find the parameters that maximize the log-likelihood.

b) **[10 points]** Express the posterior distribution of the parameters. Assume uniform priors with large enough bounds for a and $\sigma > 0$. Approximate the posterior distribution of the parameters with the Laplace approximation.

a) $y = \theta_1 x + \theta_2 \varepsilon, \ \varepsilon \sim N(0,1)$

$$P(\varepsilon) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\varepsilon^2}{2}\right)$$

Inverse Transform

$$f^{-1}(y) = \frac{y - \theta_1 x}{\theta_2} = \varepsilon$$

$$\frac{d\varepsilon}{dy} = \frac{1}{\theta_2}$$

$$P(y) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2}\left(\frac{y-\theta_1 x}{\theta_2}\right)^2\right) \frac{1}{\theta_2}$$

$$\ln P(D|\theta_1,\theta_2) = \sum_n^N \left[\ln \frac{1}{\sqrt{2\pi}} + \ln \frac{1}{\theta_2} - \frac{1}{2}\left(\frac{y_n-\theta_1 x_n}{\theta_2}\right)^2\right]$$

$$\boxed{= -\frac{1}{2} N \ln 2\pi - N \ln \theta_2 - \frac{1}{2\theta_2^2} \sum_n^N (y_n - \theta_1 x_n)^2}$$

$$\underset{\theta_1, \theta_2}{\operatorname{argmax}} \left[-\frac{1}{2} N \ln 2\pi - N \ln \theta_2 - \sum_n^N \frac{1}{2}\left(\frac{y_n-\theta_1 x_n}{\theta_2}\right)^2\right]$$

$$0 = \frac{\partial}{\partial\theta_1}\left[-\frac{1}{2} N \ln 2\pi - N \ln \theta_2 - \sum_n^N \frac{1}{2}\left(\frac{y_n-\theta_1 x_n}{\theta_2}\right)^2\right]$$

$$= -\frac{1}{2\theta_2^2} \frac{\partial}{\partial\theta_1} \sum_n^N (y_n - \theta_1 x_n)^2$$

$$= -\frac{1}{2\theta_2^2} \frac{\partial}{\partial\theta_1} \sum_n^N \left[y_n^2 - 2y_n x_n \theta_1 + x_n^2 \theta_1^2\right]$$

$$= -\frac{1}{\theta_2^2} \sum_n \left[-y_n x_n + x_n^2 \theta_1\right]$$

$$\sum_n y_n x_n = \sum_n x_n^2 \theta_1$$

$$\theta_1 = \frac{\sum_n y_n x_n}{\sum_n x_n^2} =$$

$$0 = \frac{\partial}{\partial\theta_2}\left[-\frac{1}{2} N \ln 2\pi - N \ln \theta_2 - \frac{1}{2\theta_2^2} \sum_n^N (y_n - \theta_1 x_n)^2\right]$$

$$= -\frac{N}{\theta_2} - \frac{1}{2}\sum_n^N \left[y_n^2 - 2y_n x_n \theta_1 + x_n^2 \theta_1^2\right] (-2)(\theta_2^{-3})$$

$$= \frac{-N}{\theta_2} + \frac{1}{\theta_2^3} \sum_n^N \left[y_n^2 - 2y_n x_n \theta_1 + x_n^2 \theta_1^2\right]$$

$$\frac{N}{\theta_2} = \frac{1}{\theta_2^3} \sum_n^N \left[y_n^2 - 2y_n x_n \theta_1 + x_n^2 \theta_1^2\right]$$

$$\theta_2 = \sqrt{\sum_n^N \left[y_n^2 - 2y_n x_n \theta_1 + x_n^2 \theta_1^2\right] \frac{1}{N}}$$

$$=$$

b) $P(\theta_1, \theta_2 | D) \propto P(D|\theta_1,\theta_2)\, P(\theta_1) P(\theta_2)$

$$= P(D|\theta_1,\theta_2) \frac{1}{b_1 - a_1} \frac{1}{b_2 - a_2}$$

$$P(\theta_1, \theta_2 | D) \propto P(D|\theta_1,\theta_2)$$

$$\ln P(\theta_1, \theta_2 | D)_{Lap} = -\frac{1}{2} N \ln 2\pi - N \ln \theta_{2MLE} - \frac{1}{2\theta_2^2} \sum_n^N (y_n - \theta_{1MLE} x_n)^2 + \frac{1}{2} \left[\theta_1 - \theta_{1MLE}, \theta_2 - \theta_{2MLE}\right] \begin{bmatrix} \frac{\partial^2}{\partial\theta_1^2} & \frac{\partial^2}{\partial\theta_1\theta_2} \\ \frac{\partial^2}{\partial\theta_1\theta_2} & \frac{\partial^2}{\partial\theta_2^2} \end{bmatrix} \begin{bmatrix} \theta_1 - \theta_{1MLE} \\ \theta_2 - \theta_{2MLE} \end{bmatrix}$$

$$\frac{\partial^2}{\partial\theta_1^2} = \frac{\partial}{\partial\theta_1} -\frac{1}{\theta_2^2} \sum_n \left[-y_n x_n + x_n^2 \theta_1\right]$$

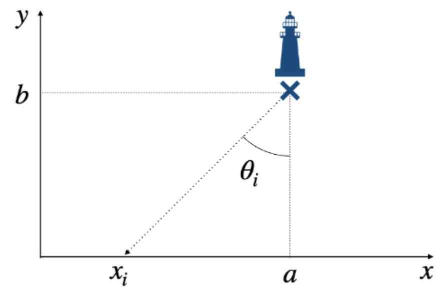
$$\boxed{= \frac{-1}{\theta_2^2} \sum x_n^2}$$

$$\frac{\partial^2}{\partial\theta_2^2} = \frac{\partial}{\partial\theta_2}\left[\frac{-N}{\theta_2} + \frac{1}{\theta_2^3} \sum_n^N \left[y_n^2 - 2y_n x_n \theta_1 + x_n^2 \theta_1^2\right]\right]$$

$$= \boxed{\frac{N}{\theta_2^2} - \frac{3}{\theta_2^4} \sum_n^N \left[y_n^2 - 2y_n x_n \theta_1 + x_n^2 \theta_1^2\right]}$$

18

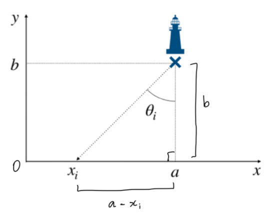
A beacon positioned at unknown location (a, b) emits light in random directions $\theta \sim \mathcal{U}(-\pi/2, \pi/2)$. The light is detected by sensors located on the shore ($y = 0$). The light detections are recorded at coordinates x_i, $i = 1, 2, \ldots, N$. Given these coordinates, where is the beacon?



a) **[10 points]** Write x_i as a function of θ_i, the angle at which the light was emitted to produce x_i. Derive the likelihood of the data x_i, $i = 1, 2, \ldots, N$ given the location of the data. Hint: x_i is the transformation of the random variable θ_i.

b) **[10 points]** Write down the posterior distribution of (a, b). Choose appropriate priors for the parameters.

a)



$$\tan\theta_i = \frac{a - x_i}{b}$$

$$\boxed{x_i = a - b\tan\theta_i}$$

$$x = a - b\tan\theta, \quad \theta \sim U(-\tfrac{\pi}{2}, \tfrac{\pi}{2})$$

$$p(\theta) = \frac{1}{\frac{\pi}{2} - -\frac{\pi}{2}} = \frac{1}{\pi}$$

Inverse Transform again

$$f^{-1}(x) = \operatorname{atan}\left(\frac{a-x}{b}\right) = \theta$$

$$\frac{d\theta}{dx} = \frac{-b}{(a-x)^2 + b^2}, \quad b \geq 0$$

$$p(\theta)\left|\frac{d\theta}{dx}\right| = \frac{1}{\pi}\frac{b}{(a-x)^2+b^2}$$

$$\boxed{P(D \mid a, b) = \frac{b^N}{\pi^N}\prod_{i=1}^{N}\frac{1}{(a-x_i)^2+b^2}}$$

b) $P(a, b \mid D) \propto P(D \mid a, b)\, P(a)\, P(b)$

$$\boxed{P(a) = U(-20, 20) = \frac{1}{40}}$$

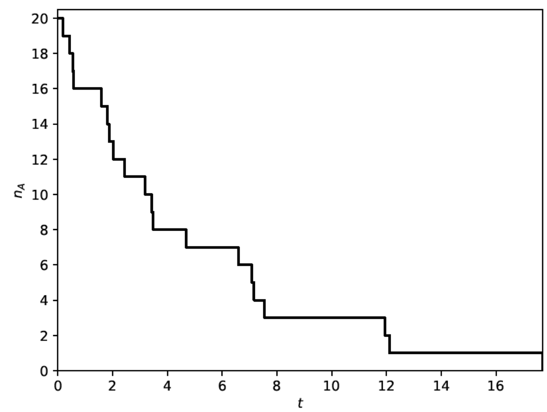
$$\boxed{P(b) = U(0, 40) = \frac{1}{40}}$$

$$P(a, b \mid D) \propto \frac{b^N}{\pi^N}\prod_{i=1}^{N}\frac{1}{(a-x_i)^2+b^2}\,\frac{1}{40}\,\frac{1}{40}$$

$$\boxed{P(a, b \mid D) = \frac{b^N}{\pi^N}\prod_{i=1}^{N}\frac{1}{(a-x_i)^2+b^2}\, C}$$

19

A decay process, $A \xrightarrow{\nu} \emptyset$, produced the following curve over time:



where n_A is the number of species A and t is the time. The values of each event t_i and the corresponding value of n_A are reported in the file `data.csv`. The process is modeled with SSA, i.e. the time between two events is sampled from an exponential distribution with a given propensity. We want to infer the value of ν.

a) Write down the probability of observing $\tau_i = t_{i+1} - t_i$ as a function of ν and $n_{A,i}$. From this, derive the likelihood of the whole dataset $D = \{t_i, n_{A,i}\}_{i=1}^{N}$ given ν.

b) Assuming a prior $p(\nu) = \lambda e^{-\lambda\nu}$ with $\lambda = 2$, write down the log-posterior density of ν. Derive analytically the maximum a posteriori estimate (MAP) of ν (the value that maximizes the posterior) and write down the Laplace approximation of the posterior. Report the numerical value of the MAP and the standard deviation σ of the Laplace approximation density.

a) $\tau = -\ln(u)\,\frac{1}{a_0}$ $\quad u \sim U(0,1)$ $\quad a_0 = n_{A_i}\, v$

$P(u) = \frac{1}{1-0}$

$u = \exp(-\tau a_0)$

$\frac{du}{d\tau} = -a_0 \exp(-\tau a_0)$

$$\boxed{P(\tau) = -n_{A_i} v \exp(-\tau n_{A_i} v)}$$

$$\boxed{P(D|v) = \prod_i^N P(t_i | n_{A_i}, v) = \prod_i^N -n_{A_i} v \exp(-(t_{i+1} - t_i) n_{A_i} v)}$$

b) $$\boxed{\ln P(v|D) = \sum_i^N \ln(-n_{A_i}) + N \ln v - (t_{i+1} - t_i) n_{A_i} v + \ln 2 - 2v}$$

$\frac{\partial}{\partial v} = \frac{N}{v} - (t_{i+1} - t_i) n_{A_i} - 2 = 0$

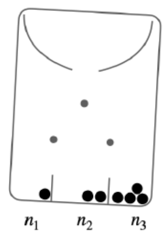
$$\boxed{v_{MAP} = \frac{N}{(t_{i+1} - t_i) n_{A_i}}}$$

$\frac{\partial^2}{\partial \theta^2} = \frac{-N}{v^2}$ always neg ✓

$$\ln P(v|D) \approx C \left[-\left(v - \frac{N}{(t_{i+1}-t_i) n_{A_i}}\right)^2 \right]$$

20

You see a tilted Galton board with two vertical rows. Beads are dropped from the top and bounce either left with probability q or right with probability $1-q$. After two bounces the beads are collected into three bins. A sketch of the experiment is shown below:



The number of beads per bin is $d = \{1,2,4\}$. We want to infer the value of of q.

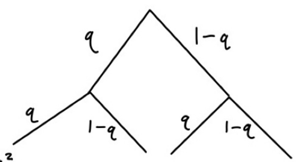
a) Write down the probability of observing the outcome $\{n_1, n_2, n_3\}$ as a function of q. From this, derive the likelihood $\ell(q\,|d)$ of the data d.

b) Given that the Galton board is tilted, assume a biased prior $\pi(q) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{B(\alpha,\beta)}$ with $\alpha = 2$ and $\beta = 6$. Write down the log-posterior density of q. $B(\alpha,\beta)$ is a normalization constant.

c) Calculate the (MAP) maximum a posterior of q analytically and verify your result using CMA-ES by directly maximizing $p(q\,|d)$.

d) Use the Laplace approximation to approximate the posterior $p(q\,|d)$. Report the value of the standard deviation σ of the Gaussian.

a)



$$q^2 + 2q(1-q) + (1-q)^2 = 1$$

$$\boxed{P(n_1\ n_2\ n_3) = q^{2n_1}\ (2q(1-q))^{n_2}\ (1-q)^{2n_3}}$$

$$\boxed{P(D|q) = \prod_i^N q^{2n_{1i}}\ (2q(1-q))^{n_{2i}}\ (1-q)^{2n_{3i}}}$$

b) $$\ln P(q\,|\,D) = 2n_{1i}\ln q + n_{2i}\ln 2 + n_{2i}\ln q + n_{2i}\ln(1-q)$$

$$+ 2n_{3i}\ln(1-q) + (\alpha-1)\ln q + (\beta-1)\ln(1-q) - \ln B$$

$$= \boxed{(2n_{1i}+n_{2i}+1)\ln q + (n_{2i}+2n_{3i}+5)\ln(1-q) + n_{2i}\ln 2 - \ln B}$$

c) $$\frac{d}{dq} = 0 = \frac{(2n_{1i}+n_{2i}+1)}{q} - \frac{(n_{2i}+2n_{3i}+5)}{1-q}$$

$$(1-q)(2n_{1i}+n_{2i}+1) = q(n_{2i}+2n_{3i}+5)$$

$$\frac{(2n_{1i}+n_{2i}+1)}{(2n_{1i}+2n_{2i}+2n_3+6)} = \boxed{q_{MAP} = \frac{1}{4}}$$

d) $$\frac{d^2}{dq^2} = \frac{-(2n_{1i}+n_{2i}+1)}{q^2} + \frac{(n_{2i}-2n_{3i}+5)}{(1-q)^2} = \frac{5}{0.25^2} + \frac{-1}{0.75^2}$$

21

The delay of the T at Harvard station is uniformly distributed in an interval $[0, \theta]$. We want to find the parameter θ by taking N measurements of the delay $x_1, x_2, ..., x_N$.
Assume, the A-priori distribution for θ is a uniform distribution in the interval [0,30] (minutes). Moreover, the likelihood is given:

$$f(x_1, x_2, ..., x_n|\theta) = \begin{cases} 0 & \theta < \text{max}(x_1, x_2, ..., x_N) \\ \frac{C}{\theta^N} & \text{max}(x_1, x_2, ..., x_N) \leq \theta \end{cases} \tag{1}$$

Here, C is an unknown normalizing constant.

a) **[8 points]** We are measuring the two delays $x_1 = 15$ (minutes) and $x_2 = 7$ (minutes). Calculate the A-posterior-density for θ based on these measurements.

b) **[6 points]** Calculate the Maximum-A-Posteriori estimate for θ.

c) **[6 points]** Calculate the probability that the parameter θ is larger than 15 (minutes) and smaller than 20 (minutes) after you have seen the two measurements.

$$P(D|\theta) = \begin{cases} 0 & \theta < \max(15,7) \\ \frac{C}{\theta^N} & \max(15,7) \leq \theta \end{cases} = \begin{cases} 0 & \theta < 15 \\ C\frac{1}{\theta^2} & 15 \leq \theta \end{cases}$$

a) $P(\theta) = \frac{1}{30-0} = \frac{1}{30}$

$$P(\theta|D) \propto P(D|\theta)\, P(\theta) = \begin{cases} 0 & \theta < 15 \\ C\frac{1}{\theta^2}\frac{1}{30} & 15 \leq \theta \leq 30 \end{cases}$$

$$C^{-1} = \int_{15}^{30} \frac{1}{30\theta^2}\, d\theta = \frac{-1}{30}\theta^{-1}\Big|_{15}^{30}$$

$$= -\frac{1}{30}\left[30^{-1} - 15^{-1}\right] = \frac{-1}{30}\left[\frac{-1}{30}\right] = \frac{1}{900}$$

$$\boxed{P(\theta|D) = 30\frac{1}{\theta^2} \quad 15 \leq \theta \leq 30}$$

b) $\max P(\theta|D) = \max 30\frac{1}{\theta^2} \quad 15 \leq \theta \leq 30$

so $P(\theta|D)$ decreases as θ increases

$\max P(\theta|D)$ is @ $\theta_{MAP} = 15$

c) $\int_{15}^{\infty} P(\theta|D)\, d\theta = 1$ so 100%

$$\int_{-\infty}^{20} P(\theta|D)\, d\theta = 30\int_{15}^{20} \frac{1}{\theta^2}\, d\theta$$

$$= -30\left[20^{-1} - 15^{-1}\right] = 0.5 \quad \text{so } 50\%$$