# Homework 1 - solutions
## Physics 589 - Fall 2018

Mikhail Tikhonov

**PROBLEM 1**

**Show that two random variables are independent if and only if for any $y_0$, the probability distribution of $x$ conditioned on $y = y_0$ is the same as the marginal $P(x)$:**

$$\forall y_0 \in \mathcal{A}_\mathcal{Y} \quad P(x|y = y_0) = P(x).$$

**Solution:**

If $x$ and $y$ are independent, then

$$P(x|y) \equiv \frac{P(x,y)}{P_Y(y)} = \frac{P_X(x)P_Y(y)}{P_Y(y)} = P_X(x).$$

Conversely, if $P(x|y) = P_X(x)$, then

$$P(x,y) = P(x|y)P_Y(y) = P_X(x)P_Y(y),$$

i.e. $x$ and $y$ are independent.

**PROBLEM 2**

**You meet Fred, who tells you he has two brothers, Alex and Bob.**

1. **What is the probability that Fred is older than Bob?**
2. **Fred adds that he's older than Alex. Conditioned on this knowledge, what is the probability that Fred is older than Bob?**

**Solution:**

Denote the three brothers $A$, $B$ and $F$. Arranging their ages in increasing order, there are 6 possibilities:

$$ABF, \quad AFB, \quad BAF, \quad BFA, \quad FAB, \quad FBA$$

Initially, all six are equiprobable, because in the absence of other information, "Alex", "Bob" and "Fred" are interchangeable labels. Unsurprisingly, the probability that Fred is older than Bob is 1/2, as this is true in three of the six equiprobable scenarios.

However, once we learn that Fred is older than Alex, only three scenarios remain: $ABF$, $AFB$, and $BAF$. Therefore, conditioned on this knowledge, the probability that Fred is older than Bob becomes 2/3.

**PROBLEM 3 (BINOMIAL STATISTICS)**

**An urn contains $K$ balls, $B$ of which are black and $W = K - B$ are white. Zoe draws a ball at random and replaces it, repeating $N$ times.**

1. **What is the probability distribution of the number of times a black ball is drawn, $n_B$?**
2. **Compute the mean $\mu_N$ and standard deviation $\sigma_N$ of $n_B$. Let $K = 20$, and $B = 5$. Compute the ratio $\sigma/\mu$ for $N = 5$ and 1000 draws.**

**Solution:**

Denote $f_B \equiv B/K$. The probability than exactly $n_B$ out of $N$ draws resulted in a black ball is given by:

$$p(n_B|f_B, N) = \binom{N}{n_B} f_B^{n_B} (1 - f_B)^{N-n_B},$$

because of the $2^N$ possible black/white ball sequences, the number of those with exactly $n_B$ blacks is given by the binomial coefficient $\binom{N}{n_B}$, and each of these $\binom{N}{n_B}$ sequences occurs with probability $f_B^{n_B} (1 - f_B)^{N-n_B}$.

To compute the mean and variance of $n_B$, we note that it is a sum over $N$ independent occurrences. For a single draw ($N = 1$), the "number of black balls" is either 1 with probability $f_B$, or 0 with probability $1 - f_B$, with an average of

$$\mu_1 = 1 \times f_B + 0 \times (1 - f_B) = f_B.$$

Similarly, the variance (mean square of deviation from the mean) for $N = 1$ is given by:

$$(\sigma_1)^2) = (1 - \mu_1)^2 \times f_B + (0 - \mu_1)^2 \times (1 - f_B) = f_B(1 - f_B)$$

For $N$ independent events, means and variances add, and we find:

$$\mu_N = N\mu_1 = N f_B \qquad \sigma_N = \sqrt{(\sigma_N)^2} = \sqrt{N(\sigma_1)^2} = \sqrt{N\, f_B\, (1 - f_B)}$$

As a result, the ratio $\sigma_N/\mu_N$ scales as $1/\sqrt{N}$:

$$\Rightarrow \qquad \frac{\sigma_N}{\mu_N} = \sqrt{\frac{1 - f_B}{N f_B}}$$

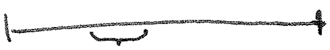
In particular, for $B = 5$ black balls out of $K = 20$, the standard-deviation-over-mean ratio is approximately 0.77 at $N = 5$, and goes down to about 0.05 at $N = 1000$.

## (4) Poisson statistics.

An event occurs stochastically with a constant average rate $r$.
Events are random and uncorrelated

(a) Probability of 0 events in time $T$?

Let's split $T$ into a large number of very short intervals



$\Delta t = T/N$

$N$ intervals of length $\Delta t = \frac{T}{N}$

For each interval, probability (no event) $= 1 - r\Delta t$

No event during $T$ means all $N$ intervals have to be event-less.

The probability of this is $\prod_{i=1}^{N} (1 - r\Delta t) = \left(1 - \frac{rT}{N}\right)^N \to \underline{\underline{e^{-rT}}}$

(as $N \to \infty$)

(b) Using the same setup ($N$ small intervals):

in each interval the number of events is at most 1 ($\Delta t$ small)
and occurs with probability $r\Delta t = \frac{rT}{N}$ $\Rightarrow$ can use the previous problem (the binomial distribution with $f = \frac{rT}{N}$).

$$p(n_T) = \binom{N}{n_T} f^{n_T} (1-f)^{N-n_T}$$

$$= \frac{N!}{n_T!\,(N-n_T)!} \left(\frac{rT}{N}\right)^{n_T} \left(1 - \frac{rT}{N}\right)^{N-n_T}$$

$$= \frac{N(N-1)\cdots(N-n_T+1)\cdot(N-n_T)!}{n_T!\,(N-n_T)!} \frac{(rT/N)^{n_T}}{\left(1-\frac{rT}{N}\right)^{n_T}} \left(1 - \frac{rT}{N}\right)^N$$

$$= \left[\frac{\frac{N}{N}\cdot\frac{N-1}{N}\cdots\frac{N-n_T+1}{N}}{\left(1-\frac{rT}{N}\right)^{n_T}}\right] \cdot \frac{(rT)^{n_T}}{n_T!} \left(1 - \frac{rT}{N}\right)^N \xrightarrow[N\to\infty]{} [1]\cdot \frac{\lambda^{n_T}}{n_T!} e^{-\lambda}$$

where $\lambda \equiv rT$

(c) The mean and standard deviation can be computed directly, but it is much easier to again re-use the results of the previous problem. The Poisson distribution is the limit of the Binomial distribution with $f = \frac{rT}{N} \equiv \frac{\lambda}{N}$, where we send $N \to \infty$ while keeping $\lambda$ constant. ⇒ Its mean and std. can be obtained in the same way.

For the binomial distribution

$$\mu = Nf$$
$$\sigma = \sqrt{Nf(1-f)}$$

Setting $f = \frac{\lambda}{N}$ and sending $N \to \infty$:

$$\mu = \lambda$$
$$\sigma = \sqrt{N \cdot \frac{\lambda}{N}\left(1 - \frac{\lambda}{N}\right)} \rightsquigarrow \sqrt{\lambda}.$$

## (5) A simple inference problem.

Bayes theorem: $$p(\text{die}\,|\,\text{data}) = \frac{p(\text{data}\,|\,\text{die}) \cdot p(\text{die})}{p(\text{data})}$$

$p(\text{die})$ ← 1/3, same for all the dice

$p(\text{data})$ ← a normalization factor, same for all dice

Let's compute $p(\text{data}\,|\,\text{die})$:

$$p(\text{data}\,|\,\text{die A}) = \frac{1}{20} \cdot \frac{4}{20} \cdot \frac{2}{20} \cdot \frac{4}{20} \cdot \frac{2}{20} \cdot \frac{4}{20} = \frac{256}{20^6} = 256\, c_0$$

$$p(\text{data}\,|\,\text{die B}) = \frac{2}{20} \cdot \frac{3}{20} \cdot \frac{2}{20} \cdot \frac{3}{20} \cdot \frac{2}{20} \cdot \frac{3}{20} = \frac{216}{20^6} = 216\, c_0$$

$$p(\text{data}\,|\,\text{die C}) = \left(\frac{2}{20}\right)^6 = \frac{64}{20^6} = 64\, c_0,$$ where $c_0$ is again the same for all the dice.

$$\Rightarrow p(\text{die A}\,|\,\text{data}) = \frac{256}{256 + 216 + 64} = \frac{256}{536} \approx 0.48$$

$$p(\text{die B}\,|\,\text{data}) = \frac{216}{536} \approx 0.40$$

$$p(\text{die C}\,|\,\text{data}) = \frac{64}{536} \approx 0.12$$

## 6 Buses in Poissonville

(a) Recalling problem 4a, the probability that time $T$ elapses with no buses is $e^{-rT}$ (where $r$ is the bus arrival rate). Let's be careful here: what this statement means is that if I pick some interval of length $T$ and count the number of buses that arrived during it, with probability $e^{-rT}$ that number is zero.

So what is the probability that, once Sally arrives at the bus stop, she has to wait for $t$ till the first bus?

$$P(\text{first bus arrives within } [t, t+\Delta t]) = \underbrace{e^{-rt}}_{\text{no buses before } t} \cdot \underbrace{r\Delta t}_{\text{a bus arrives!}}$$

So the wait time till the bus arrives is a random ~~var iable~~ variable described by probability density function

$$p(t) = re^{-rt} \quad \leftarrow \text{the exponential distribution.}$$

Its average? $\displaystyle\int_0^\infty t\, p(t)\, dt = \int_0^\infty rt\, e^{-rt}\, dt$

(integrating by parts) $\displaystyle = rt\, \frac{e^{-rt}}{-r}\Big|_0^\infty - \int_0^\infty r\, \frac{e^{-rt}}{-r}\, dt$

$$= \int_0^\infty e^{-rt}\, dt = \frac{e^{-rt}}{-r}\Big|_0^\infty = \frac{1}{r} \quad \text{(as expected).}$$

In our case, no matter when Sally arrives, her average wait time is 5 minutes

Since the problem is symmetric under $t \Leftrightarrow -t$ (uncorrelated events look the same if we record the arrival times and "run the tape backwards"), the bus Sally just missed was, on average, also 5 minutes ago.

(b) 5 mins + 5 mins = 10 mins.

(c) The average time between two consecutive buses is 5 minutes. Yet the time between the buses "just before" and "just after" Sally's arrival is 10 minutes. This seems paradoxical at first – but if Sally arrives randomly, her arrival is more likely to hit a <u>longer</u> inter-bus interval. If two ~~consecutive~~ particular buses are separated by just 1 second, this affects the mean inter-bus interval, but Sally is very unlikely to arrive just between them. So the delay between buses, as measured by Sally, is longer.

(d) See an example simulation code. (next page).

## Contents


- Buses in Poissonville: simulation
- Make plots


## Buses in Poissonville: simulation

```
clear all

meanWait = 5;
eventsN = 10000;

% Generate a randm set of bus arrival times
T = meanWait*eventsN;
bus_arrival_time=sort(T*rand(eventsN,1));

% Mean wait?
% Choose a random timepoint when "Sally" arrives at the bus stop.
salliesN = 100;
sally_t = T*rand(salliesN,1);

% To avoid "boundary effects" (e.g. the last Sally arrives after ALL buses):
% Move first bus to 0 and last one to T so there is always a previous and
% next bus. This won't skew statistics if eventsN is large
bus_arrival_time(1) = 0;
bus_arrival_time(end) = T;

% For each try, pick the closest bus before and after Sally's arrival at the bus stop
% Record the corresponding time intervals
time_to_next = NaN(salliesN,1);
time_from_prev = NaN(salliesN,1);

for k=1:salliesN
    dt = bus_arrival_time-sally_t(k);

    after = dt>0;
    time_to_next(k) = min(dt(after));

    before = dt<0;
    time_from_prev(k) = min(-dt(before));
end

% Mean interval:
fprintf('Mean time between buses: %f\n\n', mean(diff(bus_arrival_time)));

fprintf('Sally''s observations:\n')
fprintf('Mean time till next: %f\n', mean(time_to_next));
fprintf('Mean time since previous: %f\n', mean(time_from_prev));
fprintf('Mean from previous to next: %f\n', mean(time_to_next+time_from_prev));
```

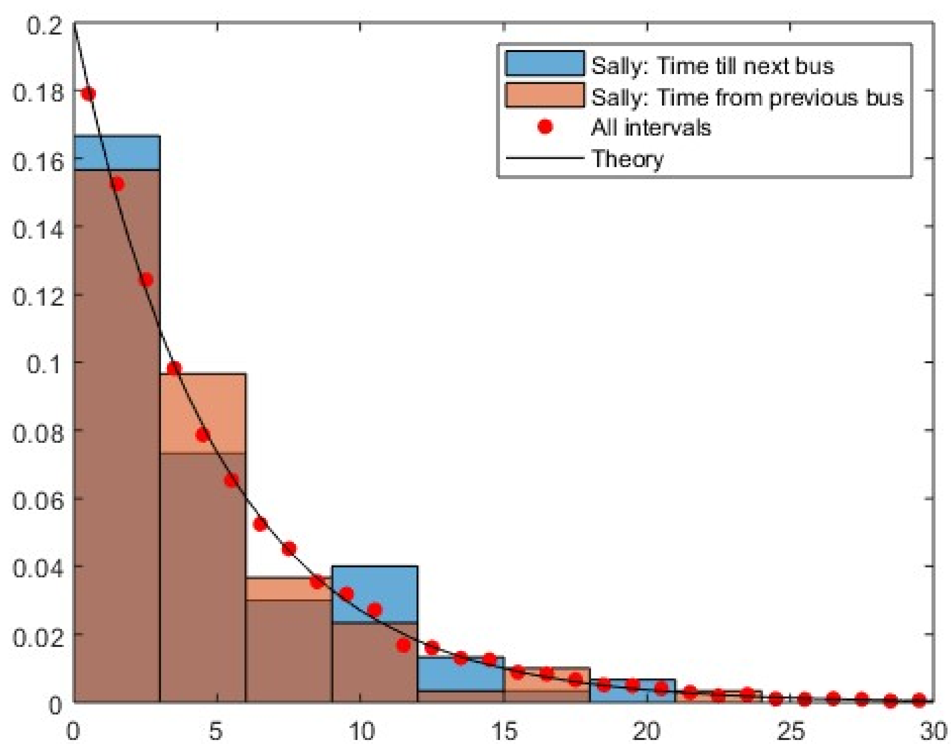
```
Mean time between buses: 5.000500

Sally's observations:
Mean time till next: 4.778824
Mean time since previous: 4.522743
Mean from previous to next: 9.301567
```

## Make plots

```
clf
histogram(time_to_next,'Normalization','pdf');
hold all
histogram(time_from_prev,'Normalization','pdf')

[cts, edges] = histcounts(diff(bus_arrival_time),'Normalization','pdf');
binCenters = edges(1:end-1)+diff(edges)/2;
plot(binCenters, cts, 'r.','MarkerSize',20);
xs = 0:0.01:30;
plot(xs, exp(-xs/meanWait)/meanWait, 'k-');
axis([0 30 0 0.2]);
legend({'Sally: Time till next bus', 'Sally: Time from previous bus', 'All intervals', 'Theory'}
)
```



## The USB problem.

Inference task: What is the probability that the cable orientation I am currently trying is correct?
"p(correct)"

Data: "What I tried so far, failed."

(a) This is an inverse problem $\Rightarrow$ Need Bayes theorem.

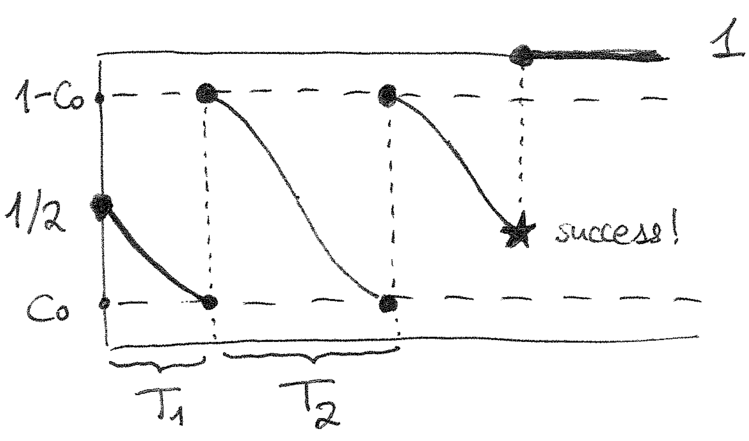
Forward probabilities: $p(\text{fail}_t \mid \text{incorrect}) = 1$
$P(\text{fail}_t \mid \text{correct}) = e^{-st}$

(because the mental computation uses the assumed rate of success $s$)

$$\Rightarrow P(\text{correct} \mid \text{fail}_t) = \frac{P(\text{fail}_t \mid \text{correct}) \cdot P(\text{correct})}{P(\text{fail}_t \mid \text{correct}) P(\text{correct}) + P(\text{fail}_t \mid \text{incorr}) P(\text{incorr})}$$

$$= \frac{e^{-st} \cdot 1/2}{e^{-st} \cdot 1/2 + 1 \cdot 1/2} = \frac{1}{1+e^{+st}}$$

(b)



(c) $$\left.\frac{1}{1+e^{st}}\right|_{t=T_1} = C_0 \Rightarrow \boxed{T_1 = \frac{1}{s}\ln\frac{1-C_0}{C_0}.}$$

Useful for later: $\frac{C_0}{1-C_0} = e^{-sT_1}$

(d) Denote $t' \equiv t - T_1$ the time elapsed after the switch.

The calculation goes just like in (a):

$$P(\text{correct} \mid \text{fail}_{t'}) = \frac{P(\text{fail}_{t'} \mid \text{correct}) \cdot P(\text{correct})}{P(\text{fail}_{t'} \mid \text{correct}) \cdot P(\text{correct}) + P(\text{fail}_{t'} \mid \text{incorr.}) P(\text{incorr.})}$$

the current, i.e. flipped orientation

elapsed since the flip

$P(\text{fail}_{t'} \mid \text{correct})$ is again $e^{-st'}$, but the "prior beliefs" (at $t'=0$) are now modified:

$$\left.\begin{aligned} P(\text{correct}) &= 1 - C_0 \\ P(\text{incorrect}) &= C_0 \end{aligned}\right\} \text{ our beliefs just after the flip.}$$

All in all: $$P(\text{correct} \mid \text{fail}_{t'}) = \frac{e^{-st'}(1-C_0)}{e^{-st'}(1-C_0) + 1 \cdot C_0} = \frac{e^{-st'}}{e^{-st'} + \frac{C_0}{1-C_0}}$$

$$= \frac{e^{-st'}}{e^{-st'} + e^{-sT_1}}.$$ Note that at $t' = T_1$, this is $1/2$ - we tried both orientations for an equal amount of time, neither succeeded so far, so we are back where we started with no reason to believe in one orientation over the other $\Rightarrow$ the problem

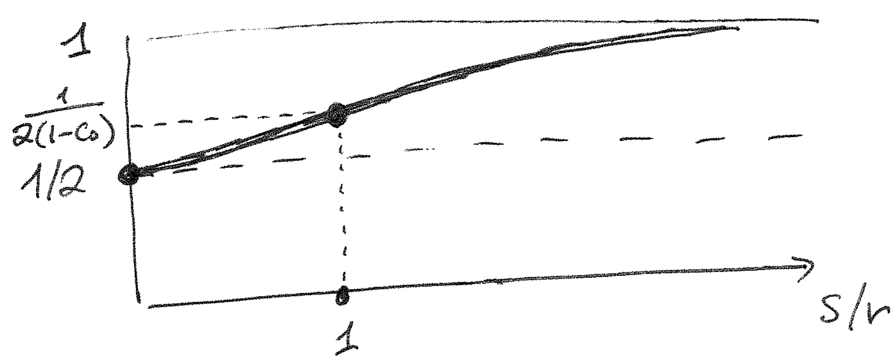
will now repeat itself. We conclude that $T_2 = 2T_1$, and (e) $T_2 = T_3 = T_4 = \ldots = 2T_1$ as well.

(f). $P(\#\text{ flips} > 0) = ?$

$$P(\#\text{flips} > 0) = P(\#\text{flips} > 0 \mid \overset{\text{initial guess}}{\text{correct}}) \cdot P(\text{correct})$$
$$+ P(\#\text{flips} > 0 \mid \text{incorrect})\, P(\text{incorrect})$$

$$= e^{-rT_1} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2} + \frac{1}{2}\left(\frac{C_0}{1-C_0}\right)^{r/s}$$

success rate $r$
trying for $T_1$
Probability of failure
(= 0 "success events")




There are two parameters under my control (part of my "strategy"): $s$ and $C_0$. But when I flip and how long it will take me to succeed depend only on their combination

$T_1 = \frac{1}{s} \ln \frac{1-C_0}{C_0}$ (and on $r$, which is what it is and is beyond my control)

From now on, we set $s = r$.

<u>Assume</u>: initial orientation guess is correct.

(g) Probability to flip the cable at least once

$\Leftrightarrow$ Probability to fail in time $T_1$, even though the guess was correct. $\Rightarrow$ $e^{-rT_1} = \frac{c_0}{1-c_0}$ (recall that $s=r$ now)

(h) Average time to success:

How much time did I spend trying? (the right way up)

If I counted only those ~~times~~ periods of time (had a stopwatch that was only running during periods $T_1$, $T_3$ etc, until the cable finally fit), the average time recorded by that stopwatch would be $1/r$. But depending on the number of flips, I pay an extra penalty of wasted time: $T_2$ (if applicable), $T_4$ (if applicable) etc.

$$\Rightarrow \text{Time to success} = 1/r + P(\#\text{flips} \geq 2 \mid \text{correct}) \times T_2 + P(\#\text{flips} \geq 4 \mid \text{correct}) \times T_4 + \ldots$$

average; assuming initial guess correct

<u>Now assume</u>: initial guess was <u>incorrect</u>.

(i) $P(\#\text{flips} \geq 1 \mid \text{incorrect}) = 1$. $P(\#\text{flips} = 1) = 1 - e^{-rT_2} = 1 - \left(\frac{c_0}{1-c_0}\right)^2$

(j) is exactly analogous to (h) above.

(k) No double flipping:

If I want to avoid double-flipping, I need to ensure that if the initial guess is correct, I get it right the first time $\Rightarrow$ $T_1$ must be long enough that probability of failing during that time is small. (Note that $T_2 = 2T_1$, so it's only the first phase that I need to worry about). longer!

$$e^{-rT_1} \leq \varepsilon \Rightarrow T_1 \geq T_1^* = -\frac{\ln \varepsilon}{r} = \frac{1}{r} \ln \frac{1}{\varepsilon}$$

Time to success? If initial guess correct: $1/r$

If initial guess incorrect: $1/r + T_1$ the "wasted" time

Average: $1/r + T_1/2 \geq 1/r + T_1^*/2 = \underline{\frac{1}{r}\left(1 + \frac{1}{2}\ln\frac{1}{\varepsilon}\right)}$

(l) If I allow double-flipping but "forbid" triple-flipping, want probability $< \varepsilon$

then $\underline{\underline{T_2}}$ must be long enough. $\Rightarrow$ $e^{-rT_2} = e^{-2rT_1} \leq \varepsilon$

$$T_1 \geq T_1^{**} = \frac{1}{2r} \ln \frac{1}{\varepsilon}$$

Time to success: If initial guess correct $1/r + T_2$ probability (flip) $= \frac{1}{r} + 2T_1 e^{-rT_1}$ see (g)

If initial guess incorrect: $1/r + T_1$

Average: $\frac{1}{2}\left(\frac{1}{r} + T_1\right) + \frac{1}{2}\left(\frac{1}{r} + 2T_1 e^{-rT_1}\right) \geq \frac{1}{r} + \frac{T_1^{**}\left(1 + 2e^{-rT_1^{**}}\right)}{2}$

$= \frac{1}{r}\left(1 + \frac{1}{2}\left(\ln\frac{1}{\varepsilon}\right)\left(\frac{1}{2} + \sqrt{\varepsilon}\right)\right)$ which is $\underline{\text{lower}}$, because $\frac{1}{2} + \sqrt{\varepsilon} < 1$.