

# Homework #5

Winter 2020, STATS 509

Dino Bektesevic

## Problem 1

A researcher is studying a certain population of animals. The researcher is interested in the proportion  $\theta$  of females. Her prior is  $\text{Beta}(\alpha, \beta)$ , where  $\alpha, \beta > 0$  are positive integers,

$$p(\theta) = \begin{cases} \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} \theta^{\alpha-1} (1-\theta)^{\beta-1} & 0 < \theta < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The researcher conducts two studies. In the first study she obtains a random sample of size  $n$  in total, that contains  $f$  females and  $m$  males.

- (a) Find the density for the researcher's posterior distribution  $p(\theta | f, m)$  for  $\theta$  after conducting the first study.

*Hint: you may cite results from lecture notes here.* From class we have the posterior given by

$$p(\theta | f, m) \approx \text{Beta}(\alpha + f, \beta + m)$$

In the second study, the researcher randomly samples animals until she observes the first male, at which point she stops. Let  $t$  be the total number of animals (both male and female) sampled in the second study.

- (b) Write down the likelihood  $p(t | \theta)$ .

$$L(\theta | t) = (1 - \theta)\theta^{t-1}$$

- (c) Using the posterior resulting from the first study (found in (a)) as the researcher's prior before seeing the data from the second study, write down an expression that is proportional to the researcher's posterior density  $p(\theta | t, f, m)$  after observing the data from **both** studies.

$$\begin{aligned} p(\theta | t, f, m) &= p(\theta)p(t | \theta) \\ &= \frac{(\alpha + f + \beta + m - 1)!}{(\alpha + f - 1)!(\beta + m - 1)!} \theta^{\alpha+f-2} (1 - \theta)^{\beta+m} \\ &\approx \theta^{\alpha+f-2} (1 - \theta)^{\beta+m} \end{aligned}$$

- (d) Using your answer to (c), or otherwise, find the researcher's posterior density,  $p(\theta | t, m, f)$  after both studies. **Hint:** Recall that for  $r, s$  positive integers,

$$\int_0^1 u^{r-1} (1-u)^{s-1} du = \frac{(r-1)!(s-1)!}{(r+s-1)!}.$$

$$\begin{aligned} 1 &= \int f(p|x) dx \\ &= c \int_0^1 \theta^{\alpha+f+t-2} (1-\theta)^{\beta+m} d\theta \\ &= c \frac{(\alpha + f + t - 2)!(\beta + m)!}{(\alpha + f + t + \beta + m - 1)!} \\ c &= \frac{(\alpha + f + t + \beta + m - 1)!}{(\alpha + f + t - 2)!(\beta + m)!} \end{aligned}$$

- (e) Using the definition of expectation, or otherwise, find  $E[\theta | t, m, f]$ , the mean of the researcher's posterior distribution.

$$\begin{aligned}
 E[\theta | t, m, f] &= \int_0^1 \theta f(\theta | t, m, f) d\theta \\
 &= \frac{(\alpha + f + t + \beta + m - 1)!}{(\alpha + f + t - 2)!(\beta + m)!} \int_0^1 \theta^{\alpha + f + t - 1} (1 - \theta)^{\beta + m} d\theta \\
 &= \frac{(\alpha + f + t + \beta + m - 1)!}{(\alpha + f + t - 2)!(\beta + m)!} \frac{(\alpha + f + t - 1)!(\beta + m)!}{(\alpha + f + t + \beta + m)} \\
 &= \frac{(\alpha + f + t + \beta + m - 1)!}{(\alpha + f + t - 2)!} \frac{(\alpha + f + t - 1)!}{(\alpha + f + t + \beta + m)} \\
 &= \frac{\alpha + f + t - 1}{\alpha + f + t + \beta + m}
 \end{aligned}$$

- (f) Suppose that another researcher had the same prior,  $\text{Beta}(\alpha, \beta)$ , but then learned about the results of both studies  $(t, m, f)$  simultaneously. Show that the posterior which results will be the same as in (d).

*Hint: note that because the studies are independent (given  $\theta$ ),  $p(t, f, m | \theta) = p(t | \theta)p(f, m | \theta)$ .*

## Problem 2

Suppose that we have a  $\text{Beta}(1, 1)$  prior for  $p$ , the mean of a Bernoulli random variable.

- (a) Find the density for  $\psi = p(1 - p)$ , the variance of a Bernoulli random variable. (Hint: use the CDF method.) Plot this density.

$$\begin{aligned}
 P(\psi \leq t) &= P(p(1 - p) \leq t) \\
 &= P(p - p^2 \leq t) \\
 &= P(-p^2 + p - t \leq 0) \\
 &= P\left(p \leq \frac{1}{2} - \sqrt{\frac{1}{4} - t}\right) + P\left(p \geq \frac{1}{2} + \sqrt{\frac{1}{4} - t}\right)
 \end{aligned}$$

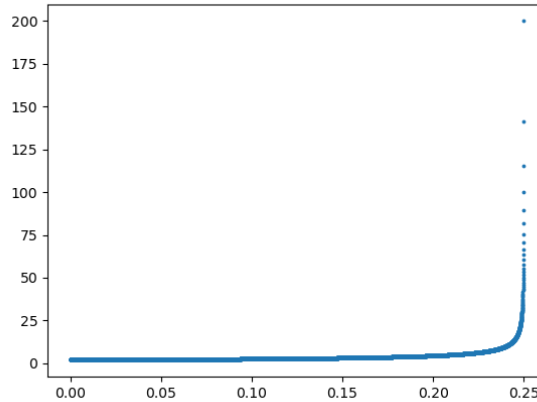
The above expression are essentially CDFs and since  $p \approx \text{Beta}(1, 1)$ , which is uniform over  $[0, 1]$ , we can just say that

$$P(\psi \leq t) = 1 - 2\sqrt{\frac{1}{4} - t}$$

The pdf then is:

$$\begin{aligned}
 f(\psi = t) &= \frac{\partial}{\partial t} F(t) \\
 &= \left(\frac{1}{4} - t\right)^{-\frac{1}{2}} \\
 &= \frac{1}{\sqrt{\frac{1}{4} - t}}
 \end{aligned}$$

for  $t \in [0, \frac{1}{4}]$



- (b) The  $\text{Beta}(1, 1)$  density for  $p$  has sometimes been claimed to be a ‘non-informative’ prior, i.e. representing a state of belief corresponding to knowing ‘nothing’. Using the plot from (a) comment briefly on this claim.

Non-informative priors don’t exactly correspond to knowing “nothing”, but instead to the belief that  $p$  is equally probable to be any value from 0 to 1. Given this assumption we can see from the plot that variance of  $V(\theta)$  is not uniform too, but that, given that assumption, more likely to be larger than, let’s say, 0.15.

### Problem 3

A researcher plans to record the number of mini-earthquake events  $X$  occurring in a given time period near a volcano. She believes this is modelled well by a Poisson distribution with parameter  $\lambda$ :

$$p(x | \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \lambda > 0.$$

Before seeing any data, her beliefs about  $\lambda$  are described by an exponential(2) distribution which has the following pdf:

$$p(\lambda) = \begin{cases} 2e^{-2\lambda} & \text{for } \lambda > 0 \\ 0 & \text{otherwise,} \end{cases}$$

- (a) Before observing the data, with what probability does the researcher believe that  $\lambda < 1$ ?

This is just purposefully confusingly written because in text they give  $\lambda = 2$  but overload the  $\lambda$  with the Poisson distribution parameter one. The answer is just the CDF of the exponential prior where, not the  $\lambda$  in the expression, is 2 and the  $\lambda$  in the expression is 1. So:

$$F(\lambda = 1) = 1 - e^{-2\lambda} = 1 - e^{-2} = 0.865$$

- (b) If the researcher observes  $x_0$  earthquakes, write down an expression that is *proportional to* the pdf  $p(\lambda | x_0)$  describing her beliefs about  $\lambda$  after making this observation.

$$\begin{aligned} p(\lambda | x_0) &= p(\lambda)p(x_0 | \lambda) \\ &= 2e^{-2\lambda} \frac{\lambda^{x_0} e^{-\lambda}}{x_0!} \\ &\approx \lambda^{x_0} e^{-3\lambda} \end{aligned}$$

- (c) Find the constant of proportionality in (b), and hence find the pdf for the researcher's posterior distribution  $p(\lambda | x_0)$ . *Hint: Note that  $\int_0^\infty t^n e^{-at} dt = n!/a^{n+1}$  for  $a > 0$ , and  $n$  a positive integer.*

$$\begin{aligned} 1 &= C \int \lambda^{x_0} e^{-3\lambda} \\ &= C \frac{x_0!}{3^{x_0+1}} \\ C &= \frac{3^{x_0+1}}{x_0!} \end{aligned}$$

- (d) Suppose that no earthquakes occur, and so  $x_0 = 0$ . Use your answer in (c) to find the probability with which the researcher now believes that  $\lambda < 1$ . What do you notice?

$$p(\lambda | x_0 = 0) = 3e^{-3\lambda}$$

Keeping in mind the confusion about  $x$  vs  $\lambda$  in the notation, we see this is an exponential distribution with the parameter  $\lambda = 3$  so evaluating a CDF for less than or equal to 1 occurrence gets us:

$$F(\lambda = 1) = 1 - e^{-3\lambda} = 1 - e^{-3} = 0.95$$

a rather high probability of one or less events.

## Problem 4

Suppose that we have a sample of size  $n$  from a Normal distribution with mean  $\mu$ , variance  $\sigma^2$ , where  $\sigma^2$  is known. The observed mean for the sample is  $\bar{x}$ .

- (a) Write down the density for  $f(\bar{x} \mid \mu)$ . This is the likelihood for  $\bar{X}$  given  $\mu$  (see note below). (*Hint: see Goldberger Section 8.3*)

Fact (Useful for part (b)) Note that if

$$h_1(\theta) = \frac{1}{a^2}(u - \theta)^2 \quad h_2(\theta) = \frac{1}{b^2}(v - \theta)^2$$

and  $h(\theta) = h_1(\theta) + h_2(\theta)$ , then

$$h(\theta) = \left(\frac{1}{a^2} + \frac{1}{b^2}\right) \left(\frac{\frac{u}{a^2} + \frac{v}{b^2}}{\frac{1}{a^2} + \frac{1}{b^2}} - \theta\right)^2 + c$$

where  $c$  is a constant (i.e. it does not depend on  $\theta$ .)

*You do not need to prove this.*

- (b) Suppose that our prior distribution for  $\mu$  is  $N(\mu_0, \tau_0^2)$ , so

$$f(\mu) = \frac{1}{\sqrt{2\pi}\tau_0} e^{-\frac{1}{2\tau_0^2}(\mu - \mu_0)^2}$$

where  $\mu_0$  and  $\tau_0^2$  are known constants. Verify the expression given for the posterior distribution of

$$\mu \mid \bar{x}$$

in the last slide of Lecture 9.

*Hint: remember that by Bayes' rule:*

$$f(\mu \mid \bar{x}) \propto f(\bar{x} \mid \mu)f(\mu)$$

*and use the expression derived in (b). It is also sufficient to simply 'spot' the form of the density. i.e. observe that  $f(\bar{x} \mid \mu)f(\mu)$  is proportional to a given known density. Thus no integration is required to answer this question.*

**Note:** to keep the notation simple above I have suppressed conditioning on the known population variance  $\sigma^2$  e.g. the likelihood could be written as:  $p(\bar{x} \mid \mu, \sigma^2)$ , the prior on  $\mu$  would then also be written as  $p(\mu \mid \sigma^2)$ . A strict Bayesian might prefer to condition explicitly on the constants to make clear what was being assumed known.

*This is a philosophical/ notational issue that is not directly relevant to the calculation you are asked to make in the question above.*

## Bayesian inference for a Normal mean with unknown variance and independent priors

Read the following notes before attempting Qu. 5:

<https://www.stat.washington.edu/tsr/s509/lecture9f02/notes-on-normal-mean-variance.pdf>

<https://www.stat.washington.edu/tsr/s509/lecture9f02/Bayesian-mean-var-for-normal-with-R.pdf>

*(For people who would like to use Python, it should be feasible to implement this using the two notes above as a guide.)*

### Problem 5

On three days between July 24 and September 5, 1882 the physicist Simon Newcomb measured the amount of time it took light to travel from his lab, to a mirror on the Washington Monument, and back to his lab. The data are here:

<https://www.stat.washington.edu/tsr/s509/examples/newcomb/newcomb-light-data.csv>

Use the following lines of R to read in the data and put the observed times into a vector  $y$ :

```
newcomb <- read.table('newcomb-light-data.csv',header=TRUE,sep=",")
y <- newcomb$time
```

(The experiments were conducted over three days, recorded in `newcomb$day`; we will not use this here.)

The times are in microseconds ( $s \times 10^{-6}$ ) (millionths of a second); the distance of the path traveled is 7.44373 km.

Suppose that the observations  $Y_1, \dots, Y_n$  are i.i.d. samples from a population with  $N(\mu, \sigma^2)$ ; where  $\mu$  is the time it takes light to travel along the path and  $\sigma^2$  describes variability due to experimental error.

There had been prior experiments by Fizeau and Foucault that attempted to measure the speed of light. Suppose that Newcomb's beliefs regarding  $\mu$  were described as follows:

$$\mu \sim N(\mu_0, 1/\tau_0^2)$$

where  $\mu_0 = 30$  and  $\tau_0^2 = 0.01$ . Thus under his prior

$$P(\mu \in [30 - 1.96/\sqrt{0.01}, 30 + 1.96/\sqrt{0.01}]) = P(\mu \in [10.4, 49.6]) = 0.95.$$

Further suppose that his prior on  $\sigma^2$  was described as follows:

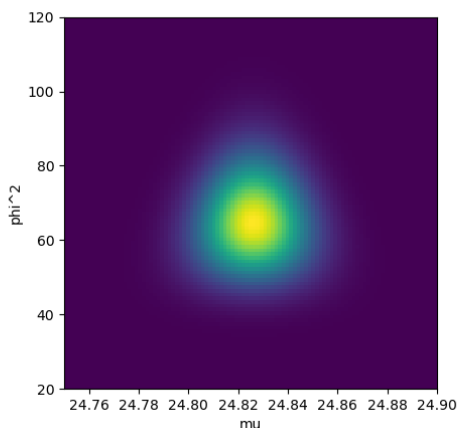
$$1/\sigma^2 = \phi^2 \sim \text{Gamma}(\nu_0/2, \nu_0/(2\phi_0^2)) = \text{Gamma}(1/2, 1/2)$$

so  $\nu_0 = 1, \phi_0^2 = 1$ . (Equivalently this is a Chi-square distribution with 1 d.f.) This prior puts high probability on  $\phi^2$  being close to 0, or equivalently, on  $\sigma^2$  being large, indicating large error.

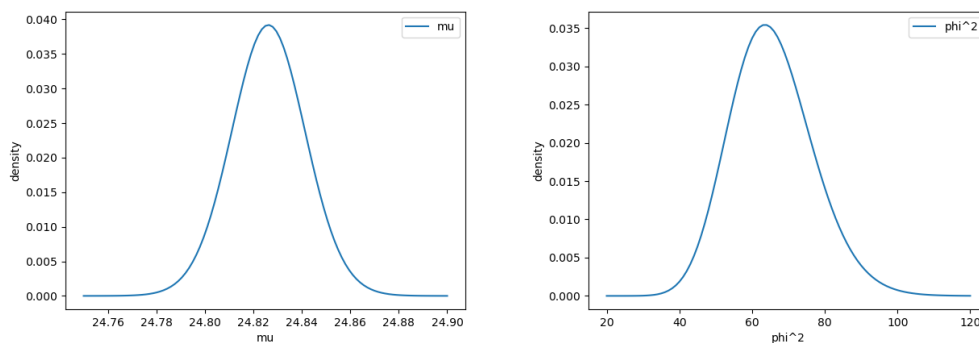
Since prior information about  $\mu$  would come from work by others on the speed of light, while information relating to  $\sigma^2$  pertains to Newcomb's set-up, which was new, it is reasonable to think that the priors on  $\mu$  and  $\sigma^2$  (or equivalently on  $\mu$  and  $\phi^2$ ) would be independent. (Note that  $\mu$  and  $\sigma^2$  are independent under the prior if and only if  $\mu$  and  $\phi^2 = 1/\sigma^2$  are independent under the prior. (Why?))

**Note:** Except for part (b)(vi), the questions below can all be answered by making minor edits to the script which will be `mean-var-for-normal.r` covered in lab on Friday.

- (a) Obtain a discrete approximation to the posterior distribution  $p(\mu, \phi^2 \mid y_1, \dots, y_n)$ . Use this to obtain:
- (i) a heat-map showing the discrete approximation to the joint posterior,



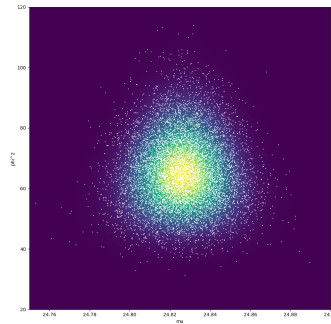
- (ii) discrete approximations to the marginal densities,  $p(\mu \mid y_1, \dots, y_n)$  and  $p(\phi^2 \mid y_1, \dots, y_n)$ ; construct plots showing these approximations.



*Hints: (1) check that the prior parameters `mu0`, `tau20`, `nu0`, `phi20` corresponding to  $\mu_0$ ,  $\tau_0^2$ ,  $\nu_0$  and  $\phi_0^2$  are set correctly. (2) To find the limits for your grid of values `mu.grid` and `phi2.grid`, it may be useful to do part (b) below first. (3) The range for `mu.grid` needs to be made quite small since the data is highly informative. (4) For the same reason the range for `phi2.grid` needs to be made large (up to 100), but should not include 0 or negative numbers as then the density is not defined.*

- (b) Use the Gibbs sampler to obtain 10,000 (dependent) draws from the posterior distribution  $p(\mu, \phi^2 \mid y_1, \dots, y_n)$ .

- (i) Construct a scatterplot showing these 10,000 draws.



- (ii) Find the 2.5%, 50% and 97.5% quantiles of the marginal posterior  $p(\mu \mid y_1, \dots, y_n)$  (by finding these quantiles in your sample of 10,000 draws).

```
mu 25th 50th and 97.5th qunatiles: [24.79595245 24.8261025 24.85651072]
```

- (iii) Find the 2.5%, 50% and 97.5% quantiles of the marginal posterior  $p(\phi^2 \mid y_1, \dots, y_n)$  (again by using your 10,000 draws).

```
phi^2 25th 50th and 97.5th qunatiles: [45.1711972 64.86764863 89.2792161 ]
```

- (iv) Find the 2.5%, 50% and 97.5% quantiles of the marginal posterior  $p(\sigma \mid y_1, \dots, y_n)$  for the population s.d.

```
sigma_phi^2 25th 50th and 97.5th qunatiles: [0.1058339 0.12416121 0.14878844]
```

- (v) As noted, the distance between Newcomb's lab and the mirror on the Washington Monument was 7.44373 km. Express the speed of light ( $c$  in km/s) as a simple function of  $\mu$ . Use this information to find the 2.5%, 50% and 97.5% quantiles of the posterior  $p(c \mid y_1, \dots, y_n)$  *Hint: this requires one line of R code using the vector of samples for  $\mu$ , which is stored in `samples[,1]`.*

```
c (km/s) quantiles: 2.994680e+05, 2.998348e+05, 3.001994e+05
```

```
From sample mean c(km/s): 2.998334970980085e-07
```

```
With sample std c(km/s): 0.0006980497158762172
```

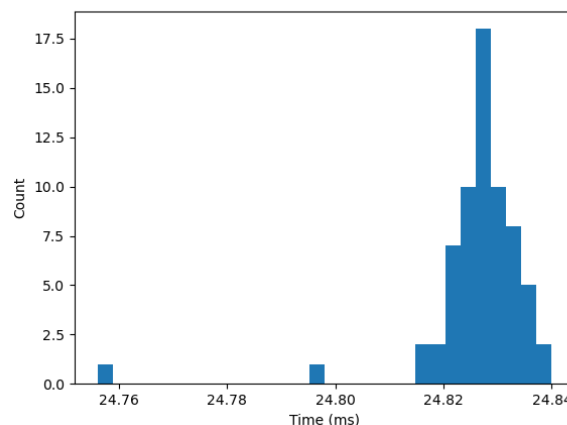
- (vi) Compare your answer in (v) to the best current measurements of the speed of light.

Pretty close to current textbook values of 2.99792458e5 km/s.

- (vii) Look at a simple histogram of the observed data; compute the sample mean and sample standard deviation. Do you see anything that might make you question the assumption that the observations follow a normal distribution?

```
Measured mean time: 24.826212121212123
```

```
Measured mean std: 0.010663610099255412
```





The two largest outliers are significantly smaller than the measured mean making the sample distribution suspicious, but given that even in these outliers are in absolute terms very close to the mean I don't think it's a problem.

*Note: the script contains various techniques for checking convergence of the Gibbs sampler, e.g. computing the auto-correlation and examining parallel boxplots. These can safely be ignored in this example!*

If you are interested, historical information on the Newcomb experiment is given [here](https://www.stat.washington.edu/tsr/s509/examples/newcomb/background-on-newcomb-experiment.pdf): `https://www.stat.washington.edu/tsr/s509/examples/newcomb/background-on-newcomb-experiment.pdf`

## Gibbs sampling

This question provides intuition as to why Gibbs sampling works. The idea is that the ‘target distribution’ that we wish to draw samples from (in Bayesian analyses this is the posterior distribution) is preserved by the steps in the procedure. (In more technical language, the procedure creates a Markov chain with a stationary distribution that is the target distribution.)

### Problem 6.

Consider the following joint density: Let  $X, Y$  be random variables with the following joint pdf:

$$f_{XY}(x, y) = \begin{cases} 2 & 0 < x < 1, \text{ and } 0 < y < x < 1 \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

- (a) Find the conditional densities  $f(x | y)$  and  $f(y | x)$ .

$$f(x|y) = \frac{f(x, y)}{f(y)}$$

$$\begin{aligned} f(y) &= \int_y^1 f(x, y) dx \\ &= \int_y^1 2 dx \\ &= 2(1 - y) \end{aligned}$$

$$f(x|y) = \frac{2}{2(1 - y)}$$

And the other way around:

$$f(x) = \int_0^x f(x, y) dy = 2x$$

$$f(x|y) = \frac{1}{x}$$

- (b) Consider the following two step Gibbs simulation scheme for finding a new point  $(X_{new}, Y_{new})$  given an existing point  $(X_{old}, Y_{old})$ :

$$\begin{aligned} X_{new} &\sim X \mid Y_{old} \\ Y_{new} &\sim Y \mid X_{new} \end{aligned}$$

where these are the conditional distributions computed in (a).

- (i) Show that if  $(X_{old}, Y_{old})$  has the joint density given by (1) then  $(X_{new}, Y_{old})$  also has the joint density (1).

*Hint: note that*

$$f(x_{new}, x_{old}, y_{old}) = f(x_{new} \mid y_{old})f(x_{old}, y_{old}),$$

*then integrate out  $x_{old}$ .*

$$\begin{aligned} f(x_n, y_o) &= \int_{y_o}^1 f(x_n \mid y_o) f(x_o, y_o) dx_o \\ &= \frac{1}{1 - y_o} \int_{y_o}^1 2 dx_o \\ &= \frac{2}{1 - y_o} (1 - y_o) \\ &= 2 \end{aligned}$$

- (ii) Similarly show that if  $(X_{old}, Y_{old})$  has the joint density given by (1) then  $(X_{new}, Y_{new})$  also has the joint density (1).

$$\begin{aligned} f(x_n, y_n) &= \int_0^{x_n} f(y_n \mid x_n) f(x_n, y_o) dy_o \\ &= \int_0^{x_n} \frac{1}{x_n} 2 dy_o \\ &= \frac{2}{x_n} \int_0^{x_n} dy_o \\ &= 2 \end{aligned}$$

(c) A friend suggests using the following sampling scheme instead:

$$\begin{aligned} X_{new}^* &\sim X \mid Y_{old} \\ Y_{new}^* &\sim Y \mid X_{old} \end{aligned}$$

Show that even if  $(X_{old}, Y_{old})$  has the joint density given by (1),  $(X_{new}^*, Y_{new}^*)$  does **not** have the density (1). *Hint: consider the support of  $(X_{new}^*, Y_{new}^*)$ . It may also help to draw a picture.*

$$\begin{aligned} f(x_n^*, y_n^*) &= \int_{y_n^*}^1 \int_0^{x_n^*} f(y_n^*, x_o, y_o) dy_o dx_o \\ &= \int_{y_n^*}^1 \int_0^{x_n^*} f(y_n^* | x_o, x_n^*, y_o) f(x_o, x_n^*, y_o) dy_o dx_o \\ &= \int_{y_n^*}^1 \int_0^{x_n^*} f(y_n^* | x_o) f(x_o, x_n^*, y_o) dy_o dx_o \\ &= \int_{y_n^*}^1 \int_0^{x_n^*} f(y_n^* | x_o) f(x_n^* | y_o) f(x_o, y_o) dy_o dx_o \\ &= \int_{y_n^*}^1 f(y_n^* | x_o) \int_0^{x_n^*} f(x_n^* | y_o) f(x_o, y_o) dy_o dx_o \\ &= \int_{y_n^*}^1 f(y_n^* | x_o) dx_o \frac{2}{x_n^*} x_n^* \\ &= 2 \int_{y_n^*}^1 \frac{1}{x_o} dx_o \\ &= -2 \ln y_n^* \end{aligned}$$

## Sampling Distributions

Questions from Goldberger:

### Problem 7

Qu. 8.3 (This is review of material from Chapters 2 and 3.)

- a. Bernouli with  $p = 0.5$

$$E[X] = 0.5$$

$$V[X] = 0.5(1 - 0.5) = 0.25$$

$$P(A) = P(0.4 \leq X \leq 0.6) =$$

- b. Normal with parameters  $\mu = 0.5$  and  $\sigma^2 = 0.25$

$$E[X] = 0.5$$

$$V[X] = 0.25$$

$$P(A) = P(0.4 \leq X \leq 0.6) = 0.159$$

- c. Exponential with  $\lambda = 2$

$$E[X] = 0.5$$

$$V[X] = 0.25$$

$$P(A) = P(0.4 \leq X \leq 0.6) = 0.148$$

## Problem 8

Qu. 8.4 (See section 8.3 on p.85.)

From section 8.3 we have that:

- a. if  $X \approx \text{Bernoulli}$ , then  $n\bar{X} \approx \text{Binomial}(n, p)$
- b. if  $X \approx \text{Normal}$ , then  $\bar{X} \approx \text{Normal}(\mu, \sigma^2/n)$
- c. if  $X \approx \text{Exponential}$ , then  $2n\lambda\bar{X} \approx \text{Chi-square}(k)$

$n=10$

- a. Bernoulli with  $p = 0.5$

$$E[\bar{X}] = 0.5$$

$$V[\bar{X}] = 0.5(1 - 0.5) = 0.025$$

$$P(A) = P(0.4 \leq X \leq 0.6) = 0.451$$

- b. Normal with parameters  $\mu = 0.5$  and  $\sigma^2 = 0.25$

$$E[\bar{X}] = 0.5$$

$$V[\bar{X}] = 0.025$$

$$P(A) = P(0.4 \leq X \leq 0.6) = 0.451$$

- c. Exponential with  $\lambda = 2$

$$E[\bar{X}] = 0.5$$

$$V[\bar{X}] = 0.25$$

$$P(A) = P(0.4 \leq X \leq 0.6) = 0.451$$

## Problem 9

Goldberger Qu. 8.7

*Hint: For part (b) read pages 87-88, Section 8.5.*

- a. The function  $Y$  is concave. So by Jensen's inequality we have

$$E[g(X)] \leq g(E[X])$$

Given is that  $E[\bar{X}] = 1/\lambda$  so  $1/E[\bar{X}] = \lambda$ . Applying Jensen's inequality to the problem then yields:

$$\lambda = \frac{1}{E[\bar{X}]} \leq E\left[\frac{1}{\bar{X}}\right] = E[T]$$

- b. If  $\lambda = 2$  and  $n = 10$  find  $E(T)$  and  $V(T)$ : From problem 8 we know that if  $X$  Exponential then  $2n\lambda\bar{X} \approx \text{Chi-Square}(2n)$  so we define  $W = 2n\lambda\bar{X} \approx \chi^2(2n)$  and look for:

$$\begin{aligned} E[1/W] &= E\left[\frac{1}{2n\lambda\bar{X}}\right] \\ &= \frac{1}{2n\lambda} E\left[\frac{1}{\bar{X}}\right] \\ &= \frac{1}{2n\lambda} E[T] \\ E[T] &= 2n\lambda E[1/W] \\ &= \frac{2n\lambda}{2n-2} \\ &= \frac{40}{18} \approx 2.222 \dots \end{aligned}$$

and the variance:

$$\begin{aligned} E[1/W^2] &= E\left[\frac{1}{4n^2\lambda^2\bar{X}^2}\right] \\ &= \frac{1}{4n^2\lambda^2} E\left[\frac{1}{\bar{X}^2}\right] \\ &= \frac{1}{4n^2\lambda^2} E[T^2] \\ E[T^2] &= 4n^2\lambda^2 E[1/W^2] \\ &= \frac{4n^2\lambda^2}{(2n-2)(2n-4)} \\ &= \frac{1600}{288} \approx 5.5555 \dots \end{aligned}$$

$$\rightarrow V[T] = E[T^2] - E[T]^2 \approx 3.3333 \dots$$

Where we used the expectation values of inverse Chi-square RV found in the book chapter 8.5 (page 88).

## Code

```
import numpy as np
from scipy import stats
from scipy.stats import norm, gamma, chi2
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

np.random.seed(0)

def problem2a():
    """Plots the PDF

    f(t) = 1/sqrt(1/4-t)

    from problem 2a.
    """
    x = np.linspace(0, 0.25, 10000)
    y = 1 / ( np.sqrt(0.25 - x) )
    plt.scatter(x, y, s=3)
    plt.show()

def readNewcombData(filepath="../../HW6Problems/newcomb-light-data.csv"):\
    """Reads supplied Newcombs data and returns read columns.
    """
    a = np.genfromtxt(filepath, delimiter=",", skip_header=1)
    return a[:,0], a[:,1]

def calcPosterior(data, mu0, tau20, nu0=1, phi20=1, muGridRange=None,
                  phi2GridRange=None, muGrid=None, phi2Grid=None, nSamples=101):\
    """Calculates posterior given in problem 5a by evaluating it over a grid.

    Parameters
    -----
    data : 'array'
        Data (Newcombs observed times)
    mu0 : 'float'
        Mean of Newcombs normal prior for the mean
    tau20 : 'float'
        STD of Newcombs normal prior for the mean
    nu0 : 'float', optional
        Nu of Newcombs gamma prior on sigma. Defaults to 1.
    phi20: 'float', optional
        Phi^2 of Newcombs gamma prior on sigma. Defaults to 1.
    muGridRange: 'tuple' or None, optional
        If given places 'nSamples' number of samples on a grid defined by
        min(muGridRange), max(muGridRange). Note either the range or the grid
        itself need to be provided.
    phi2GridRange: 'tuple', or None, optional
        If given places 'nSamples' number of samples on a grid defined by
        min(phi2GridRange), max(phi2GridRange). Note either the range or the
        grid itself need to be provided.
    muGrid : 'array' or 'None', optional
        Sampling grid.
    phi2Grid : 'array' or 'None', optional
        Sampling grid.
    nSamples : 'int', optional
        Number of samples in case only grid ranges are provided. Defaults to 101

    Returns
    -----
    posterior : 'array'
        Grid sampled posterior
    """
    G, H = nSamples, nSamples

    if muGrid is not None:
        muGrid = muGrid
    elif muGridRange is not None:
        muGrid = np.linspace(*muGridRange, num=G)
    else:
        raise ValueError("Supply mu grid or grid range.")

    if phi2Grid is not None:
        phi2Grid = phi2Grid
    elif phi2GridRange is not None:
        phi2Grid = np.linspace(*phi2GridRange, num=H)
    else:
        raise ValueError("Supply phi^2 grid or grid range.")

    tau0 = np.sqrt(tau20)
    phi0 = np.sqrt(phi20)
    phiGrid = np.sqrt(phi2Grid)

    posterior = np.zeros((G, H))
    for g in range(1, G):
        for h in range(1, H):
            a = norm.pdf(muGrid[g], loc=mu0, scale=1/tau0)
            b = gamma.pdf(phi2Grid[h], a=nu0/2, scale=(2*phi20)/nu0) #or chi2.pdf(phi2Grid[h], df=1)
            c = np.prod(norm.pdf(data, loc=muGrid[g], scale=1/phiGrid[h]))
            posterior[g,h] = a * b * c

    posterior = posterior/posterior.sum()

    return posterior

def plotPosterior(ax, posterior, xRange, yRange, xlabel="mu", ylabel="phi^2"):\
    """Given an axis, posterior and physical ranges of the posterior plots the
    posterior stretched by its physical axes.
    """
    ax.imshow(posterior.T, origin="lower", extent=(xRange, yRange))
    ax.set_xlabel(xlabel)
    ax.set_ylabel(ylabel)
    ax.set_xlim(*xRange)
```



```

ax.set_ylim(*yRange)
ax.set_aspect(1./ax.get_data_ratio())
return ax

def plotMarginal(ax, posterior, xRange, axis=0, label=''):
    """Given posterior and a physical range calculates the marginal
    of the posterior for a given axis and plots it.

    Parameters
    -----
    ax : 'matplotlib.pyplot.Axes'
        Plotting axis, (unfortunate naming convention)
    posterior : 'array'
        Posterior discretized on a grid.
    xRange : 'array'
        Range over which marginalization will happen
    axis : 'int', optional
        Axis over which to marginalize the posterior, 0 - row-wise, 1-column-wise
    label : 'str', optional
        Plot label.
    """
    marginal = posterior.sum(axis=axis)
    x = np.linspace(*xRange, num=len(marginal))
    ax.plot(x, marginal, label=label)
    ax.set_xlabel(label)
    ax.set_ylabel("density")
    ax.legend()
    return ax

def problem5a():
    """Calculates the discretized posterior from problem 5. Plots the posterior
    and the two marginals.
    """
    day, time = readNewcombData()

    muGridRange = [24.75, 24.9]
    phi2GridRange = [20, 120]
    posterior = calcPosterior(time, 30, 0.01, muGridRange=muGridRange,
                              phi2GridRange=phi2GridRange)

    fig, ax = plt.subplots()
    ax = plotPosterior(ax, posterior, muGridRange, phi2GridRange)
    plt.show()

    mus = posterior.sum(axis=1)
    fig, ax = plt.subplots()
    ax = plotMarginal(ax, posterior, muGridRange, axis=1, label="mu")
    plt.show()

    fig, ax = plt.subplots()
    ax = plotMarginal(ax, posterior, phi2GridRange, axis=0, label="phi^2")
    plt.show()

def gibbsSampler(data, nSamples=10000, mu0=30, tau20=0.01, nu0=1, phi20=1):
    """Performs Gibbs sampling as per problem 5.

    Parameters
    -----
    data : 'array'
        Data (Newcombs observed times)
    nSamples : 'int', optional
        Number of samples. Defaults to 10000
    mu0 : 'float'
        Mean of Newcombs normal prior for the mean
    tau20 : 'float'
        STD of Newcombs normal prior for the mean
    nu0 : 'float', optional
        Nu of Newcombs gamma prior on sigma. Defaults to 1.
    phi20 : 'float', optional
        Phi^2 of Newcombs gamma prior on sigma. Defaults to 1.

    Returns
    -----
    samples : 'array'
        Samples.
    """
    mean = np.mean(data)
    var = np.var(data)
    n = len(data)

    samples = np.zeros((nSamples, 2))
    samples[0,:] = np.array([mean, 1/var])
    for s in range(1, nSamples):
        prev_phi2 = samples[s-1,1]
        mustar = (mu0 * tau20 + n * mean * prev_phi2) / (tau20 + n * prev_phi2)
        phi2star = tau20 + n * prev_phi2
        mu_new = np.random.normal(mustar, 1/np.sqrt(phi2star))

        astar = (nu0 + n)/2
        bstar = (nu0/phi20 + (n-1)*var + n*(mean - mu_new)**2)/2

        phi2new = np.random.gamma(astar, 1/bstar)
        samples[s,:] = np.array([mu_new, phi2new])

    return samples

def problem5b():
    """Produces discretized posterior, gibbs sampled posterior plots and
    statistics as instructed in problem 5b.
    """
    day, time = readNewcombData()

    mu0 = 30
    tau20 = 0.01

    G, H = 101, 101

```

```

nu0 = 1
phi20 = 1

muGridRange = [24.75, 24.9]
phi2GridRange = [20, 120]
muGrid = np.linspace(*muGridRange, num=G)
phi2Grid = np.linspace(*phi2GridRange, num=G)
posterior = calcPosterior(time, 30, 0.01, muGrid=muGrid, phi2Grid=phi2Grid)

samples = gibsSampler(time)

fig, ax = plt.subplots()
ax = plotPosterior(ax, posterior, muGridRange, phi2GridRange)
ax.scatter(samples[:,0], samples[:,1], s=1, color="white", marker='+')
plt.show()

print(f"mu 25th 50th and 97.5th quantiles: {np.percentile(samples[:,0], [2.5, 50, 97.5])}")
print(f"phi^2 25th 50th and 97.5th quantiles: {np.percentile(samples[:,1], [2.5, 50, 97.5])}")
stdev = 1/np.sqrt(samples[:,1])
print(f"sigma_phi^2 25th 50th and 97.5th quantiles: {np.percentile(stdev, [2.5, 50, 97.5])}")

d = 7.44373
speedOfLight = d / (samples[:,0]*1e-6)
speedQ25, speedQ50, speedQ975 = np.percentile(speedOfLight, [2.5, 50, 97.5])
print("c (km/s) quantiles: %.6e, %.6e, %.6e"%(speedQ25, speedQ50, speedQ975))
print(f"From sample mean c(km/s): {d / time.mean()*1e-6}")
print(f"With sample std c(km/s): {d / time.std()*1e-6}")

print(f"Measured mean time: {time.mean()}")
print(f"Measured mean std: {time.std()}")

plt.figure()
plt.hist(time, bins=30)
plt.xlabel("Time (ms)")
plt.ylabel("Count")

plt.show()

if __name__ == "__main__":
    problem2a()
    problem5a()
    problem5b()

```