MATH 343 / 643 Homework #1

Professor Adam Kapelner

Due noon February 28 on github

(this document last updated 6:24pm on Saturday 22$^{\text{nd}}$ February, 2025)

Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out." Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read as much as you can online about the topics we covered.

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]" and purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 7 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com`. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: Carlos Vega

Problem 1

These are general questions about Gibbs Sampling.

(a) [easy] Let dim $[\boldsymbol{\theta}] = p$ and assume a prior $f(\boldsymbol{\theta})$ to be continuous. Describe the steps of the systematic sweep Gibbs Sampler algorithm below that will converge to $f(\boldsymbol{\theta} \mid \boldsymbol{X})$. Label the steps that are necessary for the p dimensions separately e.g. Step 2.1, Step 2.2, ..., Step 2.p. You need to reference these step numbers later on in the problem.

1st step.

0 : Initialize $\vec{\theta}_0 = [\theta_{0,1}, \theta_{0,2}, \ldots, \theta_{0,k}]$

1 : Draw $\theta_{1,1}$ from $f(\theta_1 \mid \vec{x}, \theta_2 = \theta_{0,2}, \ldots, \theta_k = \theta_{0,k}$

2 : Draw $\theta_{1,2}$ from $f(\theta_2 \mid \vec{x}, \theta_1 = \theta_{1,1}, \ldots \theta_k = \theta_{0,k})$

⋮ Repeat

k

Draw $\theta_{1,k}$ from $f(\theta_k \mid \vec{x}, \theta_1 = \theta_{1,1}, \ldots, \theta_{k-1} = \theta_{1,k-1})$

Step k+1:

Repeat steps 1 – K

using $\vec{\theta}_0 = \vec{\theta}_1$ from steps 1 – K

Step k+2:

Repeat steps 1 – K

using $\vec{\theta}_0 = \vec{\theta}_2$ from steps 1 – K

Continue until convergence on a plane number B of iterations

(b) [easy] What are all the items you need to know in order to write the code that implements a Gibbs Sampler?

1) conditional probs for each parameter.
2) the exact parameters (θ_1, θ_2)
3) initial values for all parameters
4) # of iterations
5) burn-in period length

(c) [easy] Explain what burning of the Gibbs sample chain is and why it is necessary.

Burning in Gibbs sample chains is when certain samples are discarded to mitigate initialization bias.

(d) [easy] Explain what thinning of the chain is and why it is necessary.

thinning is only keeping every ith sample from the Chain.

It is done to reduce correlation as "near" iterations depend on each other.

(e) [easy] Pretend you are estimating $\mathbb{P}(\theta_1,\ \theta_2 \mid X)$ and the joint posterior looks like the picture below where the x axis is θ_1 and the y axis is θ_2 and darker colors indicate higher probability. Begin at $[\theta_1, \theta_2] = [0.5, 0.5]$ and simulate 5 iterations of the systematic sweep Gibbs sampling algorithm by drawing new points on the plot.

contour.png

Problem 2

Consider a count model that has many zeroes. We choose to fit it with a hurdle model

$$X_1, \ldots, X_n \overset{iid}{\sim} \begin{cases} 0 & \text{w.p. } \theta_1 \\ \text{ShiftedExtNegBinomial}(\theta_2, \theta_3, +1) & \text{w.p. } 1 - \theta_1 \end{cases}$$

where the shifted distribution is just the extended negative binomial distribution so that the probability of realizing a count of one is the probability of realizing a count of zero, the probability of realizing a count of two is the probability of realizing a count of one, etc. i.e.

$$\text{ShiftedExtNegBinomial}(\theta_2, \theta_3, +1) := p(x) = \frac{\Gamma(x_i - 1 + \theta_2)}{(x_i - 1)!\Gamma(\theta_2)}(1 - \theta_3)^{x_i - 1}\theta_3^{\theta_2}.$$

(a) [harder] What is the parameter space for all three parameters of interest? This may require looking at your MATH 340 notes.

(b) [harder] Assume a ~~flat prior $f(\theta_1, \theta_2, \theta_3) \propto 1$.~~ Find the kernel of the posterior distribution $f(\theta_1, \theta_2, \theta_3 \mid \boldsymbol{x}, n_0, n_+)$ where $\boldsymbol{x} := \{x_1, \ldots, x_n\}$, the observations. Let n_0 be the number of zeroes in the dataset and $n_+ := n - n_0$, the number > 0 in the dataset.

posterior $\propto$ likelihood

$$\text{Likelihood}_1 = \theta_1^{n_0}$$

$$\text{Likelihood}_2 = (1-\theta_1)^{n_+} \cdot \prod \frac{\Gamma(x_j - 1 + \theta_2)}{(x_i - 1)!\,\Gamma(\theta_2)}^{(x_i-1)} \theta_3^{\theta_2}$$

$$L = L_1 \cdot L_2$$
$$= \theta_1^{n_0}(1-\theta_1)^{n_+} \cdot \prod \frac{\Gamma(x_j - 1 + \theta_2)}{(x_i - 1)!\,\Gamma(\theta_2)}^{(x_i-1)} \theta_3^{\theta_2}$$

(c) [easy] Find the conditional distribution $f(\theta_1 \mid \boldsymbol{x}, n_0, n_+, \theta_2, \theta_3)$ as a brand name rv.

$$\theta_1^{n_0}(1-\theta_1)^{n_+}$$

(d) [easy] Find the kernel of the conditional distribution $f(\theta_2 \mid \boldsymbol{x}, n_0, n_+, \theta_1, \theta_3)$.

$$\prod_{i=1} \frac{\Gamma(x_1 - 1 + \theta_2)}{\Gamma(\theta_2)} \theta_3^{\theta_2}$$

(e) [easy] Is the conditional distribution $f(\theta_2 \mid \boldsymbol{x}, n_0, n_+, \theta_1, \theta_3)$ a brand name rv? Yes/no

(f) [easy] Find the conditional distribution $f(\theta_3 \mid \boldsymbol{x}, n_0, n_+, \theta_1, \theta_2)$ as a brand name rv.

$$\prod_{i=1} (1-\theta_3)^{x_i - 1} \theta_3^{\theta_2}$$

(g) [easy] Is it possible to get inference for this model using a Gibbs Sampler? Why or why not?

Yes because we can sample θ_1 directly from its beta conditional dist.

Problem 3

Consider the change point model

$$X_1, X_2, \ldots, X_{\theta_3} \overset{iid}{\sim} \mathcal{N}\left(\theta_1, \sigma_1^2\right) \text{ independent of } X_{\theta_3+1}, X_{\theta_3+2}, \ldots, X_n \overset{iid}{\sim} \mathcal{N}\left(\theta_2, \sigma_2^2\right)$$

(a) [harder] What is the parameter space for all five parameters of interest?

$$\{(\theta_1, \theta_2, \theta_3, \sigma_1^2, \sigma_2^2) : \theta_1, \theta_2 \in \mathbb{R}, \sigma_1^2, \sigma_2^2 > 0, \theta_3 \in \{1, \ldots, n-1\}\}$$

(b) [harder] Assume a flat prior $\theta_1, \theta_2, \theta_3$ and Jeffrey's prior for σ_1^2, σ_2^2 which are assumed a priori independent of one another. Find the kernel of the posterior distribution.

$$f(\theta_1, \theta_2, \theta_3, \sigma_1^2, \sigma_2^2 \mid x) \propto f(x \mid \theta_1, \theta_2, \theta_3, \sigma_1^2, \sigma_1^2) \cdot f(\theta_1, \theta_2, \theta_3, \sigma_1^2, \sigma_2^2) \propto$$

$$f(x \mid \theta_1, \theta_2, \theta_3, \sigma_1^2, \sigma_2^2) \cdot f(\theta_1) \cdot f(\theta_2) f(\theta_3) \cdot f(\sigma_1^2) \cdot f(\sigma_2^2) \propto \frac{f(x \mid \theta_1, \ldots)}{\sigma_1^2 \sigma_2^2}$$

(c) [harder] Find the kernels of all five conditional distributions. If they are proportional to a known distribution, name it.

$\theta_1 = e^{-\frac{\theta_3(\theta_1 - \bar{x}_1)^2}{2\sigma_1^2}}$ $\theta_2 : e^{-\frac{(n-\theta_3)(\theta_2 - \bar{x}_2)^2}{2\sigma_2^2}}$

$\sigma_1^2 = \sigma_1^{2\,\frac{-\theta_3}{3}} e^{\left(-\frac{S_1}{2\sigma_1^2}\right)}$

(d) [harder] Find the conditional PMF of θ_3.

$\sigma_1^{2\,-\frac{k}{2}} e^{-\left(\frac{\sum_{i=1}^{k} (x_i - \theta_1)^2}{2\sigma_1^2}\right)} \sigma_2^{2\,\frac{-n-k}{2}} e^{\left(-\frac{\sum_{i=k}^{n} (x_i - \theta_2)^2}{2\sigma_2^2}\right)}$

(e) [easy] Is it possible to get inference for this model using a Gibbs Sampler? Why or why not?

Yes, we can derive all conditional distributions for each parameters.

Problem 4

Consider the discrete mixture model:

$$X_1, \ldots, X_n \overset{iid}{\sim} \begin{cases} \text{Poisson}\,(\theta_0) & \text{w.p.}\;\; \rho \\ \text{Poisson}\,(\theta_1) & \text{w.p.}\;\; 1-\rho \end{cases}$$

(a) [harder] What is the parameter space for all three parameters of interest?

$\theta_0 > 0$ $\theta_1 > 0$ $p \in [0, 1]$

(b) [harder] Assume a flat prior one all parameters. Find the kernel of the posterior distribution.

$$f(\theta_0, \theta_1, p \mid x) \propto \prod_{i=1}^{n} \left[p \cdot e^{-\theta_0} \theta_0^{x_i} + (1-p) \cdot e^{-\theta_1} \theta_1^{x_i} \right]$$

(c) [easy] Is this proportional to any known distribution?

No

(d) [harder] Is it possible to make a Gibbs Sampler to get inference here? Why or why not.

It is possible but difficult as the conditional distributions are not namebrand dist.

(e) [harder] Let's use data augmentation. Add $I_1, \ldots, I_n$ as parameters whose parameter space is $\{0, 1\}$ where $I_i = 1$ denotes that the ith observation has membership in the Poisson (θ_0) distribution and $I_i = 0$ denotes that the ith observation has membership in the Poisson (θ_1) distribution. Now find the kernel of the posterior distribution.

(f) [harder] Find the kernels of all four conditional distributions (for $\theta_0, \theta_1, \rho, I_i$). If they are proportional to a known distribution, name it.

$\theta_0 = \theta_0^{s_0} e^{-\theta_0 n_0} \sim \text{Gamma}(s_0+1, n_0)$ $\theta_1 = \theta_1^{s_1} e^{-\theta_1 n_1} \sim \text{Gamma}(s_1+1, n_1)$

$p = \rho^{n_0}(1-\rho)^{n-n_0} \sim \text{Beta}(n_0+1, n_1+1)$

$I_i = \dfrac{e^{-\theta_1}\theta_1^{x_i}}{x_i!} \cdot (1-\rho) \sim \text{Bern}(p_i)$

(g) [easy] Is it possible to get inference for this model using a Gibbs Sampler after data augmentation? Why or why not?

Yes because using data augmentation we - derive all 4 conditiona dist.

Problem 5

These are general questions about Permutation Testing.

(a) [easy] What are the null and alternative hypotheses for a two-sample permutation test?

$H_0: S_1 = S_2$ $\quad H_a\ S_1 \neq S_2$

(b) [easy] Let n_1 and n_2 be the sample sizes from population one and population two respectively. How many possible sample "permutations" are there? I put permutations in quotes because it's not truly a "permutation" in the sense that you were taught in MATH 241.

$n_{n_1} = n_1 + n_2 n_1 = \dfrac{(n_1+n_2)!}{n_1! \cdot n_2!}$

(c) [easy] Give three examples of a test statistic to employ within the body of the loop of a permutation test.

① $|\bar{X}_1 - \bar{X}_2|$ ② $Med[X_1] - Med[X_2]$ ③ $\dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

(d) [difficult] Explain how you would calculate a p-value in a permutation test.

① Choose a test statistic, T

② Calculate the test statistic

③ Generate permutations

④ Calculate the test stat for each permutation

⑤ Determine the p-value

Problem 6

These are general questions about the Bootstrap. Assume $X_1, \ldots, X_n \overset{iid}{\sim}$ some DGP.

(a) [easy] Describe the steps in the bootstrap procedure for the estimate $\hat{\hat{\theta}} := w(x_1, \ldots, x_n)$ which estimates θ.

1) Obtain the original sample

2) Compute the original estimate

3) Generate bootstrap samples

4) Compute bootstrap estimates

5) Estimate the sampling distribution

6) Calculate bootstrap statistics of interest

(b) [easy] In what situations should the bootstrap be employed instead of other inferential procedures you learned about?

When the sampling distribution is unknown or complex.

(c) [difficult] Explain in what situations the bootstrap fails. Read online about this.

if the distribution is heavy tailed, sampling with infinate variance or very heavy tails, bootstrap estimates can be extremely variable and unreliable.

Problem 7

These are questions about parametric survival using the Weibull model i.e.

$$Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Weibull}(k, \lambda) := f(y) = k\lambda^k y^{k-1} e^{-\lambda^k y^k} \mathbb{1}_{y>0}, \quad F(y) = 1 - e^{-\lambda^k y^k}, \quad S(y) = e^{-\lambda^k y^k}$$

(a) [difficult] Assume no censoring in the data. Find closed form expressions and/or equations for the MLEs of k and λ

$$L(k, \lambda : y) = \prod_{i=1}^{n} f(y_i : k, \lambda) = \prod_{i=1}^{n} k\lambda^k y_i^{k-1} e^{-\lambda y_i^k}$$

$$\ell(k, \lambda : y) = \log(L\ k, \lambda; y)$$

$$\sum_{i=1}^{n} \log(k\lambda^k y_i^{k-1} e^{-\lambda y_i^k}) = \sum(\log k + k\log\lambda + (k-1)\log y_i - \lambda y_i^k)$$

$$= n\log k + nk\log\lambda + (k-1)\sum_{i=1}^{n}\log y_i - \lambda\sum_{i=1}^{n} y_i^k \ldots$$

idk after this

(b) [easy] Assume censoring in the data so that **c** is the binary vector that is zero when censored and one if measured. Let $\boldsymbol{y}$ be the vector of measurements or censored values if not measured. Find $\ell(k, \lambda; \boldsymbol{y}, \boldsymbol{c})$.

$$\ell(k, \lambda : y, c) = \prod_{i=1}^{n} [f(y_i)]^{c_i} [S(y_i)]^{1-c_i}$$

$$\prod_{i=1}^{n} [k\lambda^k y_i^{k-1} e^{-\lambda y_i^k}]^{c_i} [e^{-\lambda y_i^k}]^{1-c_i} = \sum c_i \log(k\lambda^k y_i^{k-1}) - \lambda\sum y_i^k$$

$$= n\log k + nk\log\lambda + (k-1)\sum c_i \log y_i - \lambda \sum y_i^k$$

Problem 8

These are questions about nonparametric survival inference.

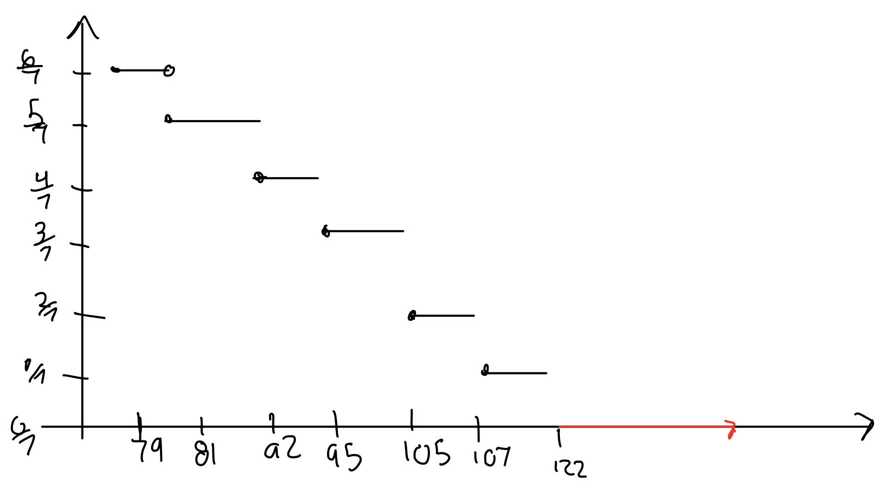
(a) [easy] Show that the empirical survival function is equal to the product limit estimator form with no censoring. Make sure to define what your notation means.

$\hat{S}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{Y_i > t}$ with/no censoring $d_i = 1$

$\hat{S}_{KM}(t) = \prod_{i : Y_i \le t}^{n} \left(1 - \frac{1}{n-i+1} \right)$

this scopes to $\frac{n-k}{n}$

(b) [easy] Consider the dataset $y = \{79, 81, 92, 95, 105, 107, 122\}$ measured in days. Draw the estimate of $S(y)$.



(c) [harder] Let your parameter of interest θ be survival past 106 days. Compute a 95% CI for θ.

$SE(\hat{S}(106)) = \frac{2}{7} \sqrt{\frac{1}{7\cdot 6} + \frac{1}{6\cdot 5} + \frac{1}{5\cdot 4} + \frac{1}{4\cdot 3} + \frac{1}{32}} \approx .1708$

$\left[\frac{2}{7} \pm 1.96 \cdot .1708 \right] = [2857 \pm .3348]$

(d) [harder] Test $H_a : \theta > 0.5$.

$$\frac{.2857 - .5}{.1708} = -1.255 < 1.645$$

retain Ha

(e) [easy] Explain how you would use the bootstrap to find a CI for the median. Explain why the bootstrap won't be so accurate in this example.

generate bootstrap samples, for each sample calculate the median, order in non-decreasing order, form 95% CI.

(f) [harder] Rederive the Kaplan-Meier estimator for the survival function.

(g) [harder] Consider the dataset $y = \{79, 81, 92+, 95, 105+, 107, 122\}$ measured in days where the "+" signs indicate censored values. Draw the Kaplan-Meier estimate of $S(y)$ in a different color atop the estimate in (b). Try to make it to scale as best as possible.

(h) [harder] Explain how you would use the bootstrap to find a CI for the median.

Generate samples, compute the K-M estimate, calculate median, construct your CI

(i) [easy] Write the hypotheses for the log-rank test.

$H_0: S_1(t) = S_2(t)$ $\qquad$ $H_a: S_1(t) \neq S_2(t)$

(j) [easy] Write the formula for the test statistic in the log-rank test.

$$Q = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$