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.

1. Initialize $\vec{\theta}_0 = [\theta_{0,1}, \theta_{0,2} \dots \theta_{0,p}]$

2.1 Draw $\theta_{1,1}$ from $f(\theta_1 \mid \vec{x}, \theta_{0,2} \dots \theta_{0,p}]$

2.2 Draw $\theta_{2,1}$ from $f(\theta_2 \mid \vec{x}, \theta_{1,1} \dots \theta_{0,p}]$

$\vdots$

2.p Draw $\theta_{k,1}$ from $f(\theta_k \mid \vec{x}, \theta_{1,1}, \dots \theta_{k-1,1})$

3 ~~8~~ Repeat step 2.1 to 2.p using $\vec{\theta}_0 = \vec{\theta}_1$

Continue until convergence or a B number of iterations

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

Need to know the distribution and the sample data, $\vec{x}$

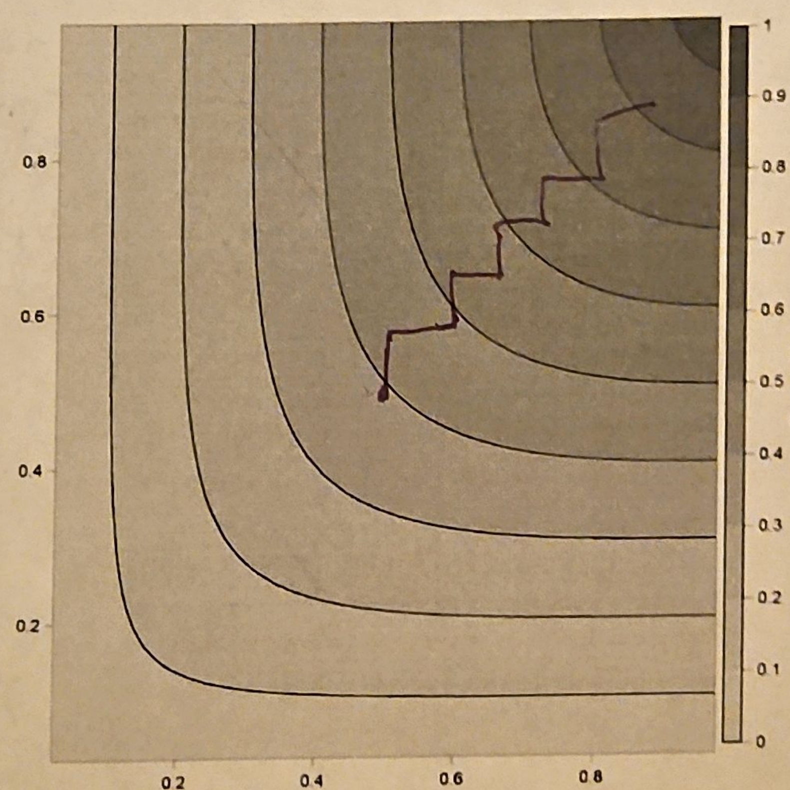
(c) [easy] Explain what burning of the Gibbs sample chain is and why it is necessary.

Burning in is eliminating the iterations that occur before convergence.
This is especially important if the initial values were poorly chosen as convergence will take a while to reflect good samples

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

Thinning the chain is ~~removal~~ when we remove some samples to ensure they stay independent. Due to Gibbs sampling method the prior sample directly effects the next one. Therefore we find a thinning value, T, and use it to thin b/w samples

(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.



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.

$\theta_1 \in [0,1]$, $\theta_2 \in \mathbb{R}$, $\theta_3 \in [0,1]$

(b) [harder] Assume a flat prior $f(\theta_1, \theta_2, \theta_3) \propto 1$. Find the kernel of the posterior distribution $f(\theta_1, \theta_2, \theta_3 \,|\, \boldsymbol{x}, n_0, n_+)$ 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.

$$f(\theta_1,\theta_2,\theta_3 \,|\, x, n_0, n_+) \propto \prod_{i=1}^{n} \theta_1^{\mathbb{1}_{x_i=0}} \left((1-\theta_1)\left(\frac{\Gamma(x_i-1+\theta_2)}{(x_i-1)!\,\Gamma(\theta_2)}(1-\theta_3)^{x_i-1}\theta_3^{\theta_2}\right)\right)^{\mathbb{1}_{x_i>0}}$$

$$= \theta_1^{n_0}(1-\theta_1)^{n_+}\frac{\Gamma(x_i-1+\theta_2)^{n_+}}{(x_i-1)!^{n_+}\,\Gamma(\theta_2)^{n_+}}(1-\theta_3)^{\sum (x_i-1)\mathbb{1}_{x_i>0}}\theta_3^{n_+\theta_2}$$

$$\propto \theta_1^{n_0}(1-\theta_1)^{n_+}\frac{\Gamma(x_i-1+\theta_2)^{n_+}}{\Gamma(\theta_2)^{n_+}}(1-\theta_3)^{\sum (x_i-1)\mathbb{1}_{x_i>0}}\theta_3^{n_+\theta_2}$$

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

$$f(\theta_1 \,|\, x, n_0, n_+, \theta_2, \theta_3) \propto \theta_1^{n_0}(1-\theta_1)^{n_+} \propto \text{Beta}(n_0+1,\, n_++1)$$

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

$$f(\theta_2 \,|\, x, n_0, n_+, \theta_1, \theta_3) \propto \frac{\Gamma(x_i-1+\theta_2)^{n_+}}{\Gamma(\theta_2)^{n_+}}\theta_3^{n_+\theta_2}$$

(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.

$$f(\theta_3 \mid x, n_0, n_+, \theta_1, \theta_2) \propto (1-\theta_3)^{\sum (X_i - 1)\mathbb{1}_{X_i > 0}} \theta_3^{n_+ \theta_2}$$

$$\propto \mathrm{Beta}(n_+\theta_2 + 1, \sum (X_i - 1)\mathbb{1}_{X_i > 0} + 1)$$

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

no because $f(\theta_2 \mid \rightarrow)$ is not known

Problem 3

Consider the change point model

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

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

$\theta_1, \theta_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$

$\theta_3 \in \mathbb{N} / \{0\}$

(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.

$$f(\theta_1, \sigma_1^2, \theta_2, \sigma_2^2, \theta_3 \mid \vec{X}) = \prod_{1}^{\theta_3} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(X_i - \theta_1)^2} \prod_{\theta_3+1}^{n} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(X_i - \theta_2)^2} \left(\frac{1}{\sigma_1^2}\right)\left(\frac{1}{\sigma_2^2}\right)$$

$$= (\sigma_1^2)^{-\theta_3/2} (\sigma_2^2)^{-(n-\theta_3)/2} e^{-(2\sigma_1^2)^{-1} \sum_1 (X_i - \theta_1)^2} e^{-(2\sigma_2^2)^{-1} \sum_{\theta_3+1} (X_i - \theta_2)^2} \sigma_1^{-2} \sigma_2^{-2}$$

$$= \sigma_1^{2(-\frac{\theta_3}{2} - 1)} (\sigma_2^2)^{(-\frac{n-\theta_3}{2} - 1)} \text{''} \qquad \text{''}$$

$$(\tfrac{n}{2} - 1) + \ldots, \sum (X_i - \theta)^2 / 2$$

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

$$f(\theta_1 \mid \theta_2, \sigma_1^2, \sigma_2^2, \theta_3) \propto N\left(\frac{\frac{\theta_3 \bar{x}}{\sigma_1^2}}{\frac{\theta_3}{\sigma_1^2}}, \frac{1}{\frac{\theta_3}{\sigma_1^2}} \right) = N\left(\frac{\theta_3 \bar{x}_1}{\theta_3}, \frac{\sigma_1^2}{\theta_3} \right)$$

$$\prod_{i=1}^{n} p$$

$$f(\theta_2 \mid \text{———}) = N\left(\bar{x}_2, \frac{\sigma_2^2}{n-\theta_3} \right)$$

$$f(\sigma_1^2 \mid \text{—}) \propto \text{Inv Gamma}\left(-\frac{\theta_3}{2} - 1, \frac{\sum_{1}^{\theta_3} (x_i - \theta_1)^2}{2} \right)$$

$$f(\sigma_2^2 \mid \text{—}) \propto \text{Inv Gamma}\left(\frac{-(n-\theta_3)}{2} - 1, \frac{\sum_{\theta_3+1}^{n} (x_i - \theta_2)^2}{2} \right)$$

$$p(\theta_3 \mid \text{—}) \propto (\sigma_1^2)^{-\frac{\theta_3}{2}} (\sigma_2^2)^{-\frac{(n-\theta_3)}{2}} e^{-\frac{1}{2\sigma_1^2} \sum_{1}^{\theta_3} (x_i - \theta)} e^{-\frac{1}{2\sigma_2^2} \sum_{\theta_3+1} (x_i - \theta_2)}$$

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

$$p(\theta_3 \mid \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \vec{x}) \propto \frac{ \sigma_1^{2^{-\frac{\theta_3}{2}}} \sigma_2^{2^{-\frac{(n-\theta_3)}{2}}} e^{-\frac{1}{2\sigma_1^2} \sum_{1}^{\theta_3} (x_i - \theta)^2} e^{-\frac{1}{2\sigma_2^2} \sum (x_i - \theta_2)^2} }{ \sum_{\theta_3 \in \{1, \ldots, n-1\}} \sigma_1^{2^{-\frac{\theta_3}{2}}} \sigma_2^{2^{-\frac{(n-\theta_3)}{2}}} e^{-\frac{1}{2\sigma_1^2} \sum_{1}^{\theta_3} (x_i - \theta)^2} e^{-\frac{1}{2\sigma_2^2} \sum (x_i - \theta_2)^2} }$$

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

No because θ_3 is not a known distribution

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?

$\theta_0 \in (0, \infty)$ $\theta_1 \in (0, \infty)$ $p \in [0, 1]$

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

$\theta_0, \theta_1, p \propto 1$

$$f(\theta_0,\theta_1,p|\vec{x}) = \prod_{i=1}^{n} p\left(\frac{e^{-\theta_0}\theta_0^{x_i}}{x_i!}\right) + (1-p)\left(\frac{e^{-\theta_1}\theta_1^{x_i}}{x_i!}\right)$$

$$\propto \prod_{i=1}^{n} p(e^{-\theta_0}\theta_0^{x_i}) + (1-p)e^{-\theta_1}\theta_1^{x_i}$$

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

No

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

No because we cannot sample from $f(\theta_0|\theta_1,p,\vec{x})$, $f(\theta_1|\theta_0,p,\vec{x})$, $f(p|\theta_0,\theta_1,p,\vec{x})$ as they are not known distributions

[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.

$$f(\theta_0,\theta_1,p,\vec{I}_i|\vec{x}) \propto \prod_{i=1}^{n} \left(p(e^{-\theta_0}\theta_0^{x_i})\right)^{I_i} \left((1-p)e^{-\theta_1}\theta_1^{x_i}\right)^{(1-I_i)}$$

$\Sigma I_i = n_0$, $\Sigma 1-I_i = n-n_0$

$$= p^{n_0}(1-p)^{n-n_0} e^{-n_0\theta_0}\theta_0^{\Sigma I_i x_i} e^{-(n-n_0)\theta_1}\theta_1^{\Sigma(1-I_i)x_i}$$

(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.

$$f(\theta_0 | \theta_1, p, \vec{x}) = e^{-n_0 \theta_0} \theta_0^{\sum I_i x_i} \propto \text{Gamma}\left(\sum I_i x_i + 1, \; -n_0 \theta_0\right)$$

$$f(\theta_1 | \theta_0, p, \vec{x}) = e^{-(n-n_0)\theta_1} \theta_1^{\sum (1-I_i) x_i} \propto \text{Gamma}\left(\sum (1-I_i) x_i + 1\right), \; -(n-n_0)\theta_1)$$

$$f(p | \theta_0, \theta_1, \vec{x}) = p^{n_0} (1-p)^{n-n_0} \propto \text{Beta}(n_0 + 1, \; n - n_0 + 1)$$

$$f(I_i | \theta_0, \theta_1, p, \vec{x}) = \theta_0^{\sum I_i x_i} \theta_1^{\sum (1-I_i) x_i} p^{\sum I_i} (1-p)^{\sum 1 - I_i} e^{-\theta_0 \sum I_i} e^{-\theta_1 \sum (1-I_i)}$$

$$\left(p \theta_0^{x_i} e^{-\theta_0}\right)^{I_i} \left((1-p) e^{-\theta_1} \theta_1^{x_i}\right)^{1-I_i}$$

$$\text{Bern}\left(\frac{p \theta_0^{x_i} e^{-\theta_0}}{p \theta_0^{x_i} e^{-\theta_0} + (1-p) e^{-\theta_1} \theta_1^{x_i}}\right)$$

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

yes because we know all distribution for priors

Problem 5

These are general questions about Permutation Testing.

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

$H_a: DGP_1 \neq DGP_2$ $\qquad$ $H_0: DGP_1 = DGP_2 = DGP$

(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.

$$\binom{n_1 + n_2}{n_1}$$

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

small sample size or samples too dependent on each other

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

$$L(k,\lambda) = \prod_{i=1}^{n} k\lambda^k y_i^{k-1} e^{-\lambda y_i^k} = k^n \lambda^{nk} \sum y_i^{k-1} e^{-\lambda \sum y_i^k}$$

$$\ell(k,\lambda) = nk + nk(\ln \lambda) + (k-1) \sum \ln(y_i) - \lambda \sum y_i^k$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{nk}{\lambda} - \sum y_i^k = 0$$

$$\hat{\lambda}^{mle} = \frac{nk}{\sum y_i^k}$$

$$\frac{d\ell}{dk} = \frac{n}{k} + n\ln(\lambda) + \sum \ln(y_i) - \lambda \sum y_i^k \ln(y_i) = 0$$

$$\frac{n}{k} - \lambda \sum y_i^k \ln(y_i) = -n\ln(\lambda) - \sum \ln(y_i)$$

(b) [easy] Assume censoring in the data so that $\mathbf{c}$ 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 $\ell(k, \lambda; \boldsymbol{y}, \mathbf{c})$.

$$\mathcal{L}(k,\lambda; \vec{y}, \vec{c}) = \prod_{\{i: c_i=1\}} f(y) \prod_{\{i: c_i=0\}} P(Y > y_i)$$

$$= \prod \lambda^k k y_i^{k-1} e^{-\lambda^k y^k} \prod e^{-\lambda^k y^k}$$

$$= (\lambda^k k)^{n_1} \prod_{\{i: c_i=1\}} y_i^{k-1} e^{-\lambda^k \sum_{c_i=1} y_i^k} e^{-\lambda^k \sum_{c_i=0} y_i^k}$$

what

$n_1 = \sum \mathbb{1}_{c_i=1}$

$n_0 = \sum \mathbb{1}_{c_i=0}$

$$\ell(k,\lambda; \vec{y}, \vec{c}) = n_1 k \ln(\lambda) + n_1 \ln(k) + (k-1) \sum_{i: c_i=1} \ln(y_i) - \lambda^k \sum_{i=1}^{n} y_i^k$$

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.

time	# of deaths	# still survivors
$t_0 = 0$	0	$n = n_1$
t_1	d_1	$n_2 = n_1 - d_1$
t_2	d_2	$n_3 = n_2 - d_2$
t_3	d_3	$n_4 = n_3 - d_3$
$\vdots$	$\vdots$	$\vdots$
t_f	d_f	0

$$S(t_3) = P(T>t_3 \mid T>t_2)\,P(T>t_2 \mid T>t_1)\,P(T>t_1)$$

$$= \frac{n_4}{n_3}\cdot\frac{n_3}{n_2}\cdot\frac{n_2}{n_1} = \frac{n_3-d_3}{n_3}\cdot\frac{n_2-d_2}{n_2}\cdot\frac{n_1-d_1}{n_1}$$

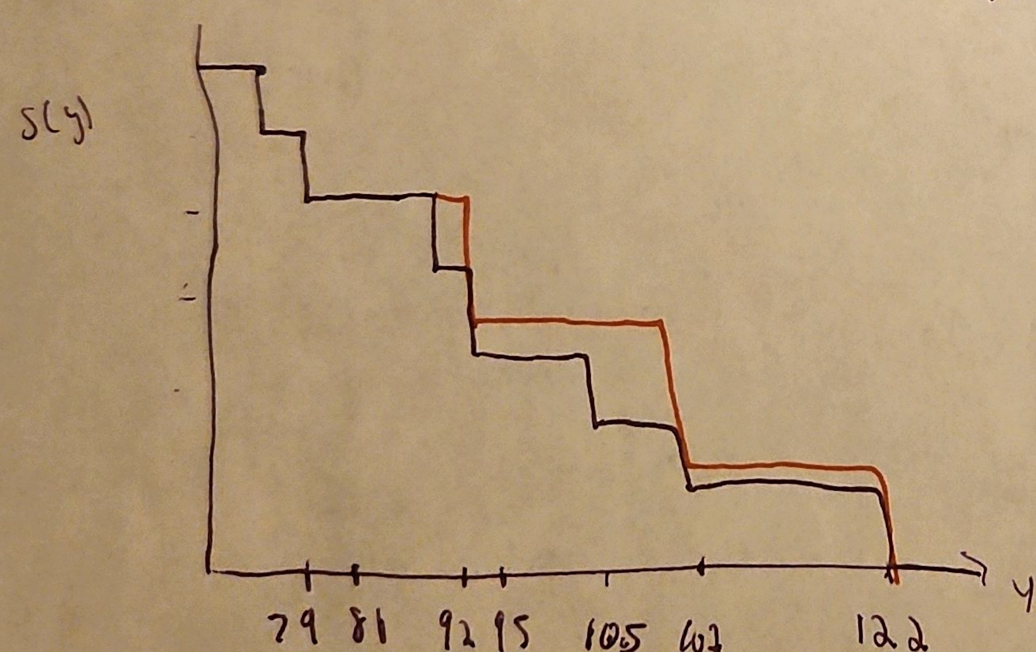
$$= \left(1-\frac{d_3}{n_3}\right)\cdot\left(1-\frac{d_2}{n_2}\right)\cdot\left(1-\frac{d_1}{n_1}\right)$$

$$S(t_k) = \prod_{i=0}^{k-1}\left(1-\frac{d_i}{n_i}\right)$$

d_i = # died in the time period

n_i = # at risk or still alive before t_k

(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.

$$S(106) \approx \frac{2}{7} = .286$$

$$CI_{95\%} = \left[.286 \pm 1.96\sqrt{\frac{(.286)(.714)}{7}}\right]$$

$$= [-.048,\ .6207]$$

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

$H_0 : \theta \leq 0.5 \in CI$

Retain H_0

(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.

Create large number of samples, from resampling the sample we have with replacement.

Then create $\hat{\theta}_{median}$ for each sample.

Create $\hat{\theta}_{boot}$ distribution from samples then find the bounds of $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ to create CI, won't be accurate because sample is small

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

time	d_i	q_i	n_i
			$n_1 = n$
$t_0 = 0$	0	0	
t_1	d_1	0	$n_2 = n_1 - d_1 - q_1$
s_1	0	q_1	N/A
t_2	d_2	0	$n_3 = n_2 - d_2$
			$n_4 = n_3 - d_3 - q_3$
t_3	d_3	0	
s_2	0	q_2	N/A

$$\hat{S}(y) = \prod \left(1 - \frac{d_i}{n_i}\right)$$

(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.

Sample the data with replacement B times and create $\hat{\theta}$ statistic

Create $\hat{\theta}_{boot}$ distribution and take $\frac{\alpha}{2}, 1-\frac{\alpha}{2}$ for CI bounds

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

$H_1: DGP_1 \neq PGP_2 \Rightarrow H_0: DGP = PGP_2$

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

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