18.650. Fundamentals of Statistics
Fall 2023. Recitation sheet 2

1 The Bootstrap Method

Problem 1 Let $X_1, \dots, X_n$ be i.i.d. $Ber(p)$ random variables. Let $X_1^*, \dots, X_m^*$ be a bootstrap sample from the X_is.

1. What is the distribution of X_1^* (not conditioned on $X_1, \dots, X_n$)?
Solution: $\mathbb{P}(X_1^* = 1) = \sum_{i=1}^n \mathbb{P}(X_i = 1) \cdot \frac{1}{n} = p$, so it is a $Ber(p)$ random variable.

2. What is the distribution of X_1^* conditioned on $X_1, \dots, X_n$? Compute the expectation and variance of X_1^* conditioned on $X_1, \dots, X_n$.
Solution: Note that X_1^* is randomly selected from the X_i. As such, the probability it equals 1 is the proportion of X_i which are 1. This is $\bar{X}_n$ so conditioned on the X_i we have $X_1^* \sim Ber(\bar{X}_n)$. Therefore, the expectation and variance are $\bar{X}_n$ and $\bar{X}_n(1-\bar{X}_n)$ respectively.
3. What is $\mathbb{P}(X_1^* = X_2^*)$ and $\mathbb{P}(X_1^* = X_2^* | X_1, \dots, X_n)$? As $n \to \infty$, what do both of these values tend towards?
Solution: For two i.i.d. $Ber(p)$ variables X_1, X_2 we have:
$$\mathbb{P}(X_1 = X_2) = \mathbb{P}(X_1 = X_2 = 1) + \mathbb{P}(X_1 = X_2 = 0) = p^2 + (1-p)^2.$$
Conditional on $X_1, \dots, X_n$, the X_i^* are i.i.d. $Ber(\bar{X}_n)$ random variables so we get:
$$\mathbb{P}(X_1^* = X_2^* | X_1, \dots, X_n) = \bar{X}_n{}^2 + (1-\bar{X}_n)^2.$$
In the unconditional case there are two possibilities: if X_1^*, X_2^* are drawn from the same X_i they are equal with probability 1. Otherwise, the two are i.i.d. $Ber(p)$ variables so the probability they are equal is the same as $\mathbb{P}(X_1 = X_2)$. As such:
$$\mathbb{P}(X_1^* = X_2^*) = \frac{1}{n}(1) + \frac{n-1}{n}(p^2 + (1-p)^2).$$
Both of these tend to $\mathbb{P}(X_1 = X_2)$ as $n \to \infty$.

Problem 2 Let $X_1, \dots, X_n$ be uniform random variables on the range $[0, \theta]$ for some unknown θ.

1. Describe how one could get an estimate on the variance of θ using the bootstrap. Use the estimator $\hat{\theta} = \max(X_1, \dots, X_n)$ and draw B samples.
   Solution: Let $X^*_{1,i}, \dots, X^*_{n,i}$ be bootstrap samples of $X_1, \dots, X_n$ for $i = 1, \dots, B$. Let
   $$T^*_{n,i} = \max(X^*_{1,i}, \dots, X^*_{n,i}).$$
   Then the bootstrap estimate for the variance is:
   $$v_{boot} = \frac{1}{B}\sum_{i=1}^{B}(T^*_{n,i} - \bar{T}^*_n)^2$$
   where $\bar{T}^*_n$ is the average of the $T^*_{n,i}$.
2. Suppose $\hat{\theta}_n = 9.79, B = 40$ and our bootstrap estimates of θ consist of 27 occurrences of 9.79, 8 occurrences of 9.75, 2 occurrences of 9.39, and 3 occurrences of 9.33. Use pivotal intervals to come up with a 90% confidence interval for θ.
   Solution: Our 5% top and bottom sample quantiles are 9.33 and 9.79 respectively. Using pivotal intervals we get
   $$(2\hat{\theta}_n - \theta^*_{.95}, 2\hat{\theta}_n - \theta^*_{.05}) = (2(9.79) - 9.79, 2(9.79) - 9.33) = (9.79, 10.25).$$

Problem 3 (AoS Exercise 8.4) Let $X_1, \dots, X_n$ be distinct observations (no ties). Show that there are $\binom{2n-1}{n}$ distinct bootstrap samples.
Solution: To find the number of distinct bootstrap samples, we simply need to partition the n bootstrap values by which copy of X_i they will sample from. Thus, the question is equivalent to finding the number of partitions $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n$ such that $\lambda_i \geq 0$ and $\sum_{i=1}^n \lambda_i = n$. We can use a simple combinatorial argument (e.g. stars and bars) to find that the number of such samples is simply $\binom{(n+n)-1}{n} = \binom{2n-1}{n}$.

Problem 4 Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} \text{Poiss}(\lambda)$, for some unknown $\lambda > 0$ and let λ_0 be a fixed (known) positive number.

1. Consider the following hypotheses:
   $$H_0 : \lambda = \lambda_0 \text{ vs. } H_1 : \lambda \neq \lambda_0.$$
   Give a test with asymptotic level 5%.
   Solution: We'd like some test statistic T and some region R (which may change with n) such that as $n \to \infty$ we have
   $$\mathbb{P}_{\lambda_0}(T \in R) \to 0.05.$$

In the terminology from class we'd then set $\Psi = \{T \in R\}$. We know that $\hat{\lambda} = \bar{X_n}$ is the MLE of λ, so let's use this to come up with the desired region. By the CLT, we know that $X_i \sim \mathsf{Poiss}(\lambda_0)$ satisfy:

$$\frac{\hat{\lambda} - \lambda_0}{\sqrt{\lambda_0/n}} \to \mathcal{N}(0,1)$$

where the denominator is $\sqrt{\lambda_0/n}$ as $Var(\bar{X_n}) = \lambda_0/n$. This tells us that asymptotically,

$$\mathbb{P}\left(\frac{|\hat{\lambda} - \lambda_0|}{\sqrt{\lambda_0/n}} \geq 1.96\right) \to 0.05.$$

Thus we can use $\hat{\lambda}$ as our test statistic T and R to be the complement of the interval $(\lambda_0 - 1.96\sqrt{\lambda_0/n}, \lambda_0 + 1.96\sqrt{\lambda_0/n})$.

2. Consider the following hypotheses:

$$H_0 : \lambda \leq \lambda_0 \text{ vs. } H_1 : \lambda > \lambda_0.$$

Give a test with asymptotic level (at most) 5%.

Solution: In this case the null hypothesis contains multiple different values of λ, so instead we're looking for a region R and test statistic T with:

$$\sup_{\lambda \leq \lambda_0} \mathbb{P}_\lambda(T \in R) \leq 0.05.$$

In other words we need the bound to hold for all $\lambda \leq \lambda_0$. There are two complications from our previous method: (1) λ can take multiple values so it's unclear whether or not we should still center R around λ_0 and (2) The variance of the true distribution changes with λ_0. We'll resolve the first issue by using a one-sided interval and the second issue by bounding the variance. By the central limit theorem:

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \to \mathcal{N}(0,1).$$

Thus,

$$\mathbb{P}\left(\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \geq 1.64\right) \to 0.05$$

which yields:

$$\mathbb{P}_\lambda(\hat{\lambda} \geq \lambda + 1.64\sqrt{\lambda/n}) \to 0.05.$$

Increasing λ will only decrease the probability of this interval. As such for all $\lambda \leq \lambda_0$:

$$\lim_{n\to\infty} \mathbb{P}_\lambda\left(\hat{\lambda} \geq \lambda_0 + 1.64\sqrt{\lambda_0/n}\right) \leq 0.05.$$

Thus we can take $T = \hat{\lambda}$ and $R = [\lambda_0 + 1.64\sqrt{\lambda_0/n}, \infty)$.

3. Consider the following hypotheses:

$$H_0 : \lambda \geq \lambda_0 \text{ vs. } H_1 : \lambda < \lambda_0.$$

Give a test with asymptotic level (at most) 5%.
Solution: Since our inequality on λ is the reverse of last time it would make sense that we'd simply want to take the reverse of our previous inequality:

$$\mathbb{P}\left(\frac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} \leq -1.64\right) \to 0.05$$

which yields:

$$\mathbb{P}_\lambda(\hat{\lambda} \leq \lambda - 1.64\sqrt{\lambda/n}) \to 0.05.$$

However this cannot be as straightforwardly bounded as λ achieves its minimum at λ_0, but $-1.64\sqrt{\lambda/n}$ achieves its maximum there. To resolve this, we can use Slutsky's theorem to replace the λ in the variance by $\bar{X}_n$ and get:

$$\mathbb{P}_\lambda\left(\hat{\lambda} \leq \lambda - 1.64\sqrt{\bar{X}_n/n}\right) \to 0.05.$$

Now we can just replace λ by its minimum λ_0 to get $T = \hat{\lambda}$ and $R = (-\infty, \lambda_0 - 1.64\sqrt{\bar{X}_n/n}]$.

4. Consider the following hypotheses:

$$H_0 : |\lambda - 2| \leq 1 \text{ vs. } H_1 : |\lambda - 2| > 1.$$

Give a test with asymptotic level (at most) 5%.
Solution: Note that in both previous cases we've ended up testing whether $\bar{X}_n \leq c_1$ and/or $\bar{X}_n \geq c_2$ for some c_1, c_2. Let's look for something similar here. We have that

$$\begin{aligned}\sup_{\lambda\in[1,3]} \mathbb{P}_\lambda(c_2 \leq \bar{X}_n \leq c_1) &\leq \sup_{\lambda\in[1,3]} \mathbb{P}_\lambda(\bar{X}_n \leq c_1) + \mathbb{P}_\lambda(\bar{X}_n \geq c_2) \\ &\leq \mathbb{P}_1(\bar{X}_n \leq c_1) + \mathbb{P}_3(\bar{X}_n \geq c_2).\end{aligned}$$

The last line here uses the fact that $\bar{X}_n$ will increase as λ increases so $\bar{X}_n \leq c_1$ will be most likely when $\lambda = 1$. This inequality tells us that if we perform both the $\lambda \geq 1$ test and $\lambda \leq 3$ test from the previous two parts at level .025 then we'll get the desired combined test.

Problem 5 (AoS Exercise 10.15) Let $X \sim \text{Bin}(n, p)$. Construct the likelihood ratio test for

$$H_0 : p = p_0 \text{ versus } H_1 : p \neq p_0.$$

Compare to the Wald test.

Solution: (Note that we will not be covering the likelihood ratio test.) If we consider the MLE $\hat{p}$, we know in this case that it is asymptotically normal. Then the Wald test for size α is simply

$$T = W = \frac{\hat{p} - p}{\hat{se}}, R = \{|W| > z_{\alpha/2}\},$$

where $\hat{se} = \sqrt{\hat{p}(1-\hat{p})/n}$.

Problem 6 Suppose we have a dataset $X_1, \ldots, X_n$ that is i.i.d. with density $f_\theta(x) = \theta x^{-\theta-1}\mathbb{1}_{x>1}$, $\theta > 1$.

1. Calculate the MLE and Fisher Information for θ.
   Solution: The log-likelihood function is

$$l(X_1, \ldots, X_n; \theta) = n \log \theta - (\theta + 1) \sum_{i=1}^{n} \log X_i$$

$$\implies l'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} \log X_i = 0$$

$$\implies \hat{\theta} = \frac{n}{\sum_{i=1}^{n} \log X_i}.$$

   Since $l''(\theta) = -\frac{n}{\theta^2} < 0$, $\hat{\theta}$ is indeed the MLE. Thus, the Fisher infermation is $I(\theta) = \mathbb{E}[-l''(X;\theta)] = \frac{1}{\theta^2}$.

2. Give the likelihood ratio test with level $\alpha = 0.05$ to test if $\theta = 2$.
   Solution: N/A.

3. Give Wald's test with level $\alpha = 0.05$ to test if $\theta = 2$.
   Solution: As before, we use the asmyptotic normality of the MLE to find that

$$\sqrt{n}\frac{\hat{\theta} - \theta}{\sqrt{I^{-1}(\hat{\theta})}} = \sqrt{n}\left(\frac{\hat{\theta} - \theta}{\hat{\theta}}\right) \rightsquigarrow \mathcal{N}(0, 1).$$

   Thus, our Wald's test is simply to reject $H_0 : \theta = 2$ if

$$|W| = \left|\sqrt{n}\left(\frac{\hat{\theta} - \theta}{\hat{\theta}}\right)\right| \geq 1.96 = z_{\alpha/2}.$$

4. Assume $n = 100$ and $\hat{\theta} = 2.45$. Compute the asymptotic p-values of both tests.
   Solution: (We won't compute the p-value for the likelihood ratio test.) The p-value of the Wald test is easy to compute: our test statistic is $W = \sqrt{100}\left(\frac{2.45-2}{2.45}\right) \approx 1.84$, giving us a p-value of 0.066.

Problem 7 We want to test if a 4-sided die is fair. To that end, we rolled the die 500 times. The following table represents both the data collected and the true proportion of rolls that should come up with each number, assuming the die is fair.

Dice Side	1	2	3	4	Total
# of rolls	134	150	109	107	500

Choose an appropriate test to verify that this 4-sided die is fair (set up hypotheses, test statistic, and compute the p-value).
Solution: Let $p = (p(1), p(2), p(3), p(4))$ denote the true pmf associated with the die, and $p_0 = (0.25, 0.25, 0.25, 0.25)$ denote the pmf being tested. We wish to test

$$H_0 : p = p_0, \quad H_1 : p \neq p_0.$$

Note that the expected rolls for each number is $0.25n$, so the test statistic is

$$T_n = \sum_{i=1}^{4} \frac{(O_i - 0.25n)^2}{0.25n}.$$

Using $n = 500$ and our observed values, we get

$$\begin{aligned} T_n = \sum_{i=1}^{4} \frac{(O_i - 125)^2}{125} &= \frac{(134-125)^2}{125} + \frac{(150-125)^2}{125} + \frac{(109-125)^2}{125} + \frac{(107-125)^2}{125} \\ &= \frac{81 + 625 + 256 + 324}{125} = \frac{1286}{125} = 10.288. \end{aligned}$$

Since $T_n \to \chi^2_3$ in distribution, we get a p-value of around 0.01627.

Problem 8 Consider a multiple hypothesis problem with $N = 6$ independent tests. We observe the following p-values:

Test 1: 0.0502
Test 2: 0.0345
Test 3: 0.0605
Test 4: 0.0225
Test 5: 0.0060
Test 6: 0.0170

1. Suppose the tests were all of level $\alpha = 5\%$. Which tests do we reject?
   Solution: We see that tests 2, 4, 5, and 6 have levels less than 0.05, so we reject tests 2, 4, 5, 6.
2. Suppose we apply the Bonferroni method to control the FWER by $\alpha = 5\%$. Which tests do we reject?
   Solution: If we apply the Bonferroni method, we will want to reject any test whose p-value is below $0.05/6 \approx 0.0083$. The only such test is test 5, so we reject test 5.
3. Suppose we apply the Benjamini-Hochberg procedure to control the FDR by $\alpha = 5\%$. Which tests do we reject?
   Solution: We order the p-values of the tests in increasing order: tests 5 ($P^{(1)}$), 6 ($P^{(2)}$), 4 ($P^{(3)}$), 2 ($P^{(4)}$), 1 ($P^{(5)}$), 3 ($P^{(6)}$). The highest i such that $P^{(i)} \leq 0.05i/6$ is $i = 3$, so we reject the tests corresponding to $P^{(1)}, P^{(2)}, P^{(3)}$, which are tests $5, 6, 4$.

Problem 9 Let $X_1, \ldots, X_n \sim Uniform(0, \theta)$ and let $Y = \max\{X_1, \ldots, X_n\}$. We want to test

$$H_0 : \theta = \frac{1}{2} \quad \text{vs.} \quad H_1 : \theta > \frac{1}{2}.$$

The Wald test is not appropriate since Y does not converge to a Normal distribution. Suppose we decide to test this hypothesis by rejecting H_0 when $Y > c$.

1. Find the power function.
   Solution: The power function for this test is

$$\beta(\theta) = \mathbb{P}_\theta(Y > c) = 1 - \mathbb{P}_\theta(Y \leq c) = 1 - \prod_{i=1}^{n} \mathbb{P}_\theta(X_i \leq \theta) = 1 - \prod_{i=1}^{n} \min(\frac{X_i}{\theta}, 1).$$

2. What choice of c will make the size of the test .05?
   Solution: The test has size α when $1 - (\min(\frac{c}{\theta}, 1))^n = \alpha$. Thus, we require $c = 1/2 * 0.95^{1/n}$.
3. In a sample of size $n = 20$ with $Y = 0.48$ what is the p-value? What conclusion about H_0 would you make?
   Solution: For $n = 20$, $Y = 0.48$, we have a p-value of $1 - (0.48/0.5)^{20} = 0.558$. This means that the test provides little or no evidence against H_0.
4. In a sample of size $n = 20$ with $Y = 0.52$ what is the p-value? What conclusion about H_0 would you make?
   Solution: For $n = 20$, $Y = 0.52$, we have a p-value of $1 - 1^{20} = 0$. We thus have extremely strong evidence against H_0.

Problem 10 In 1861, 10 essays appeared in the New Orleans Daily Crescent. They were signed "Quintus Curtuis Snodgrass" and some people suspected they were actually written by Mark Twain. To investigate this, we will investigate the proportion of three letter words found in an author's work. From eight Twain essays we have:

.225	.262	.217	.240	.230	.229	.235	.217

From 10 Snodgrass essays we have:

.209	.205	.196	.210	.202	.207	.224	.223	.220	.201

1. Perform a Wald test for equality of the means. Use the nonparametric plug-in estimator. Report the p-value for the difference of means. What do you conclude?
   Solution: We calculate $\bar{X} - \bar{Y} = 0.022$ and $\hat{se} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 0.006$. Our Wald test statistic is thus 3.704, corresponding to a p-value of 0.0002. This is very strong evidence against H_0.
2. Now use a permutation test to avoid the use of large sample methods. What is your conclusion?
   Solution: Generating permutations to create an empirical CDF, we see that our new p-value is around 0.0005, which is still very strong evidence against H_0.

2 Bayesian Inference

Problem 11 For this problem, note that a $\mathsf{Beta}(a, b)$ distribution has density $f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$ and expectation $a/(a+b)$, where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. The $\mathsf{Gamma}(a, b)$ distribution has density $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$ and mean a/b. In each case below, where the prior distribution and the sample distribution is given:

a. Find the posterior distribution in a Bayesian approach, and

b. Compute the Bayes estimator.

1. $p \sim \mathsf{Beta}(a, b)$ for some $a, b > 0$ and conditional on $p, X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Ber}(p)$.
   Solution: Note that

$$\begin{aligned} \pi(p|X_1, X_2, \ldots, X_n) &\propto \pi(p) L(X_1, X_2, \ldots, X_n|p) \\ &\propto (p^{a-1}(1-p)^{b-1})(p^{\sum X_i}(1-p)^{n-\sum X_i}) \\ &\propto p^{a+(\sum X_i)-1}(1-p)^{b+n-(\sum X_i)-1}, \end{aligned}$$

and so the posterior distribution is a $\mathsf{Beta}(a + \sum X_i, b + n - \sum X_i)$ distribution. The Bayes estimator is the expectation of this distribution, which is $\frac{a+\sum X_i}{a+b+n}$.

2. $\pi(\theta) = 1, \forall \theta > 0$ and conditional on $\theta, X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Unif}([0,\theta])$.
Solution: Note that

$$\begin{aligned} \pi(\theta | X_1, X_2, \ldots, X_n) &\propto \pi(\theta) L(X_1, X_2, \ldots, X_n | \theta) \\ &\propto \frac{1}{\theta^n} \mathbf{1}(X_1, \ldots, X_n \leq \theta) \\ &= \frac{1}{\theta^n} \mathbf{1}(X_{(n)} \leq \theta), \end{aligned}$$

where $X_{(n)}$ denotes the maximum of the X_i's, and so the posterior distribution is

$$\pi(\theta | X_1, X_2, \ldots, X_n) = \frac{C}{\theta^n} \mathbf{1}(X_{(n)} \leq \theta), C = \frac{1}{\int_{X_{(n)}}^{\infty} \frac{1}{x^n} dx} = (n-1)(X_{(n)})^{n-1}.$$

The Bayes estimator is the expectation of this distribution, which is

$$\int_{X_{(n)}}^{\infty} \theta \pi(\theta | X_1, X_2, \ldots, X_n) = \int_{X_{(n)}}^{\infty} \frac{C}{\theta^{n-1}} = \frac{C}{(n-2)(X_{(n)})^{n-2}} = \frac{(n-1)X_{(n)}}{n-2}.$$

3. $\lambda \sim \mathsf{Exp}(1/\alpha)$ for some $\alpha > 0$ and conditional on $\lambda, X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Exp}(1/\lambda)$.
Solution: Note that

$$\begin{aligned} \pi(\lambda | X_1, X_2, \ldots, X_n) &\propto \pi(\lambda) L(X_1, X_2, \ldots, X_n | \lambda) \\ &\propto (\alpha e^{-\alpha\lambda})(\lambda^n e^{-\lambda \sum X_i}) \\ &\propto \lambda^n e^{-\lambda(\alpha + \sum X_i)}, \end{aligned}$$

and so the posterior distribution is a $\mathsf{Gamma}(n+1, \alpha + \sum X_i)$ distribution. The Bayes estimator is the expectation of this distribution, which is $\frac{n+1}{\alpha + \sum X_i}$.

4. $\lambda \sim \mathsf{Exp}(1/\alpha)$ for some $\alpha > 0$ and conditional on $\lambda, X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Poiss}(\lambda)$.
Solution: Note that

$$\begin{aligned} \pi(\lambda | X_1, X_2, \ldots, X_n) &\propto \pi(\lambda) L(X_1, X_2, \ldots, X_n | \lambda) \\ &\propto (\alpha e^{-\alpha\lambda})(\lambda^{\sum X_i} e^{-\lambda n}) \\ &\propto \lambda^{\sum X_i} e^{-\lambda(\alpha + n)}, \end{aligned}$$

and so the posterior distribution is a $\mathsf{Gamma}(\sum X_i + 1, \alpha + n)$ distribution. The Bayes estimator is the expectation of this distribution, which is $\frac{\sum X_i + 1}{\alpha + n}$.

Problem 12 (AoS Exercise 11.3) Let $X_1, \dots, X_n \sim \mathsf{Unif}(0, \theta)$. Let $f(\theta) \propto 1/\theta$. Find the posterior density.

Solution: The posterior density is proportional to the likelihood times the prior:

$$\begin{aligned} f(\theta|X_1, X_2, \dots, X_n) &\propto f(\theta)\mathcal{L}(X_1, X_2, \dots, X_n|\theta) \\ &\propto (\frac{1}{\theta})\prod_{i=1}^{n}(\frac{\mathbf{1}(X_i \le \theta)}{\theta}) \\ &\propto \theta^{-(n+1)}\mathbf{1}(X_{(n)} \le \theta), \end{aligned}$$

where $X_{(n)}$ denotes the maximum of the X_i's. To normalize the density, we find the constant to be

$$C = \frac{1}{\int_{X_{(n)}}^{\infty} \frac{1}{x^{n+1}}} = n(X_{(n)})^n,$$

so our posterior density is

$$f(\theta|X_1, X_2, \dots, X_n) = \frac{n}{\theta}\left(\frac{X_{(n)}}{\theta}\right)^n \mathbf{1}(X_{(n)} \le \theta).$$

Problem 13 (AoS Exercise 11.6) Let $X_1, \dots, X_n \sim \mathsf{Poiss}(\lambda)$.

1. Let $\lambda \sim \mathsf{Gamma}(\alpha, \beta)$ be the prior. Show that the posterior is also a Gamma. Find the posterior mean.
   Solution: Note that

$$\begin{aligned} \pi(\lambda|X_1, X_2, \dots, X_n) &\propto \pi(\lambda)L(X_1, X_2, \dots, X_n|\lambda) \\ &\propto (\lambda^{\alpha-1}e^{-\beta\lambda})(\lambda^{\sum X_i}e^{-\lambda n}) \\ &\propto \lambda^{\alpha+\sum X_i - 1}e^{-\lambda(\beta+n)}, \end{aligned}$$

   and so the posterior distribution is a $\mathsf{Gamma}(\alpha + \sum X_i, \beta + n)$ distribution. Thus, the posterior mean is just $\frac{\alpha+\sum X_i}{\beta+n}$.
2. Find the Jeffreys' prior. Find the posterior.
   Solution: We saw before in a previous recitation that the Fisher information for a Poisson distribution is simply $\frac{1}{\lambda}$. The Jeffreys' prior essentially considers a prior proportional to $\sqrt{I(\theta)}$, which in this case gives $\pi(\lambda) \propto \sqrt{\frac{1}{\lambda}}$. Applying the same logic as in the first part of the problem, we see that

$$\begin{aligned} \pi(\lambda|X_1, X_2, \dots, X_n) &\propto \pi(\lambda)L(X_1, X_2, \dots, X_n|\lambda) \\ &\propto (\lambda^{-1/2})(\lambda^{\sum X_i}e^{-\lambda n}) \\ &\propto \lambda^{1/2+\sum X_i - 1}e^{-\lambda n}, \end{aligned}$$

   and so the posterior distribution is a $\mathsf{Gamma}(1/2 + \sum X_i, n)$ distribution.

Problem 14

I have five coins in my pocket. I know a priori that one gives heads with probability .2 and the other four give heads with probability .6. Assume that I pull out one of the five coins at random from my pocket (each coin has probability 1/5 of being pulled out) and I want to find out which of the two types of coin it is. To that end, I flip the coin 6 times and record the results $X_1, \ldots, X_6$ of each coin flip where $X_i = 1$ if "heads" and $X_i = 0$ if "tails". Let $p = \mathbb{P}(X_1 = 1)$.

1. What is the parameter space of possible values for p in this context?
   Solution: The parameter space is $\{0.2, 0.6\}$.
2. Write the pmf π that quantifies my prior knowledge on p.
   Solution: The pmf is
   $$\pi(0.2) = \frac{1}{5}, \pi(0.6) = \frac{4}{5},$$
   as 1 coin has probability of flipping heads equal to 0.2 while the other 4 coins have probability 0.6.

   Assume that we observe $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 2$.
3. Compute the value of the maximum likelihood estimator for this parameter space.
   Solution: Note that
   $$\mathcal{L}(X_1, \ldots, X_6 | p = 0.2) = (\frac{1}{5})^2 (\frac{4}{5})^{6-2} = \frac{256}{5^6},$$
   while
   $$\mathcal{L}(X_1, \ldots, X_6 | p = 0.6) = (\frac{3}{5})^2 (\frac{2}{5})^{6-2} = \frac{144}{5^6},$$
   so the value of p that maximizes the likelihood is $\hat{p} = 0.2$.
4. Compute the value of the Bayes estimator.
   Solution: We can calculate the ratio
   $$\frac{\pi(p = 0.2 | X_1, \ldots, X_6)}{\pi(p = 0.6 | X_1, \ldots, X_6)} = \frac{\pi(0.2)\mathcal{L}(X_1, \ldots, X_6 | p = 0.2)}{\pi(0.6)\mathcal{L}(X_1, \ldots, X_6 | p = 0.6)} = \frac{(\frac{1}{5})(\frac{256}{5^6})}{(\frac{4}{5})(\frac{144}{5^6})} = \frac{4}{9},$$
   and since $\pi(p = 0.2 | X_1, \ldots, X_6) + \pi(p = 0.6 | X_1, \ldots, X_6) = 1$, we have that $\pi(p = 0.2 | X_1, \ldots, X_6) = \frac{4}{13}$ while $\pi(p = 0.6 | X_1, \ldots, X_6) = \frac{9}{13}$. Thus, the Bayes estimator is
   $$0.2 \cdot \frac{4}{13} + 0.6 \cdot \frac{9}{13} = \frac{31}{65}.$$
5. What is the value of the maximum a posteriori (MAP) estimator?
   Solution: From the previous part, we saw that $\pi(p = 0.2 | X_1, \ldots, X_6) = \frac{4}{13}$ while $\pi(p = 0.6 | X_1, \ldots, X_6) = \frac{9}{13}$. Thus, the maximum a posteriori is 0.6 (as it maximizes $\pi(p | X_1, \ldots, X_6)$).

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

The table lists $P(Z \leq z)$ where $Z \sim N(0,1)$ for positive values of z.