Problem 3: Review through an Exponential Model

Setup

All problems on this page follow this setup:

Recall that the exponential distribution with parameter $\lambda > 0$ has density

$$g(x) = \lambda e^{-\lambda x}, \quad (x \geq 0).$$

We write $X \sim \text{Exp}(\lambda)$ when a random variable X has this distribution.

The Gamma distribution with positive parameters α (shape), β (rate) has density

$$h(x) \propto x^{\alpha-1} e^{-\beta x}, \quad (x \geq 0).$$

and has expectation $\frac{\alpha}{\beta}$. We write $X \sim \text{Gamma}(\alpha, \beta)$ when a random variable X has this distribution.

Suppose we have independent and identically distributed random variables $X_1, \ldots, X_n$, that we model as coming from an exponential distribution with some unknown parameter $\lambda > 0$. In short, $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Exp}(\lambda)$.

(a) Bayesian Estimation and Confidence Regions: Calculations

6/6 points (graded)

Compute the Jeffreys prior $\pi_J(\lambda)$. If the answer is improper, enter your answer such that $\pi_J(1) = 1$.

For $\lambda > 0$,

$\pi_J(\lambda) =$ 1/lambda ✔ **Answer:** 1/lambda

$\frac{1}{\lambda}$

In a particular scenario, we have $n = 10$ observations $X_1, \ldots, X_{10}$ with $\sum_{i=1}^{10} X_i = 12$. Recall $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Exp}(\lambda)$.

Using the Bayesian approach with the Jeffreys prior $\pi_J(\lambda)$ computed above as the prior, compute the posterior distribution $\pi(\lambda | X_1, \ldots, X_{10})$.

(Enter the posterior distribution in proportionality notation without worrying about the normalization factor.)

$\pi(\lambda | X_1, \ldots, X_{10}) \propto$

lambda^9*exp(-lambda*12) ✔ **Answer:** lambda^9*e^(-12*lambda)

Compute Bayes' estimator, which is defined in lecture as the mean of the posterior distribution.

(Enter your answer accurate to at least 3 decimal places.)

$\hat{\lambda}^{\text{Bayes}} =$ 10/12 ✔ **Answer:** 0.8333

Compute the maximum-a-posteriori (MAP) estimator.

(Enter your answer accurate to at least 3 decimal places.)

$\hat{\lambda}^{\text{MAP}} =$ 3/4 ✔ **Answer:** 0.7500 这里我好像对数求导，求的极值这样做。答案用mode做的。

Compute a *one-sided* Bayesian confidence region with level 0.05 that takes the form $[a, \infty)$. That is, if a random variable $X \sim \pi(\lambda|X_1, \ldots, X_{10})$, we are finding a such that $P(X \geq a) = 0.95$.

Similarly, compute a *one-sided* Bayesian confidence region with level 0.05 that takes the form $(0, b)$. That is, if a random variable $X \sim \pi(\lambda|X_1, \ldots, X_{10})$, we are finding b such that $P(0 < X < b) = 0.95$.

Enter the values of a and b, rounded to three decimal places.

(You would need computational tools to access the inverse cumulative distribution function of the Gamma distribution. Feel free to use R, or this Gamma distribution calculator. Beware that b on this website is the scale parameter, which is different from the rate parameter β, and $b = 1/\beta$.)

$a =$ 0.452117 ✔ **Answer:** 0.452 $b =$ 1.308768 ✔ **Answer:** 1.309

Correction Note (May 21): An earlier version of the problem statement contained an error. The second last paragraph read: "Compute a *one-sided* Bayesian confidence region with level 0.05 that takes the form $[a, \infty)$. That is, if a random variable $X \sim \pi(\lambda|X_1, \ldots, X_{10})$, we are finding a such that $P(X \geq a) = 0.05$."

Correction Note (May 23): An earlier version of the problem statement did not include the instruction "(Enter the posterior distribution in proportionality notation without worrying about the normalization factor.)"

Correction Note (June 2): An earlier version of the problem statement include the note "If the answer is improper, enter your answer such that $\pi_J(1) = 1$." but used the equal sign in the prompt.

Solution:

- **Jeffreys Prior:** Recall the definition of the Jeffreys prior to be $\pi_J(\lambda) \propto \sqrt{I(\lambda)}$ for the case where λ is a single real parameter, where $I(\lambda)$ is the Fisher information. We compute this quantity.

$$I(\lambda) = \mathbb{E}\left[\left(\frac{d}{d\lambda}\log L(X_i|\lambda)\right)^2\right],$$

and plugging in $L(X_i|\lambda) = \lambda e^{-\lambda x}$ we get

$$\log L(X_i|\lambda) = -\lambda x + \log\lambda$$

and so

$$\frac{d}{d\lambda}\log L(X_i|\lambda) = -x + \frac{1}{\lambda}.$$

Going back, we get that

$$\begin{aligned} I(\lambda) &= \mathbb{E}\left[\left(\frac{d}{d\lambda}\log L(X_i|\lambda)\right)^2\right] \\ &= \mathbb{E}\left[\left(-X + \frac{1}{\lambda}\right)^2\right] \\ &= \mathbb{E}\left[(-X)^2\right] - \frac{2}{\lambda}\mathbb{E}[X] + \left(\frac{1}{\lambda}\right)^2 \\ &= \frac{2}{\lambda^2} - \frac{2}{\lambda}\frac{1}{\lambda} + \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \end{aligned}$$

Hence,

$$\pi_J(\lambda) \propto \sqrt{I(\lambda)} = \sqrt{\frac{1}{\lambda^2}} = \boxed{\frac{1}{\lambda}}$$

- **Posterior Distribution:** As $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Exp}(\lambda)$,

$$L(X_i|\lambda) = \lambda e^{-X_i\lambda}$$

and

$$L(X_1, \ldots, X_n|\lambda) = \prod_{i=1}^{n} L(X_i|\lambda).$$

By Bayes formula, we could write

$$\begin{aligned}
\pi(\lambda|X_1, \ldots, X_n) &\propto \pi(\lambda) L(X_1, \ldots, X_n|\lambda) \\
&= \pi(\lambda) \prod_{i=1}^{n} L(X_i|\lambda) \\
&= \frac{1}{\lambda} \prod_{i=1}^{n} \lambda e^{-X_i\lambda} \\
&= \frac{1}{\lambda} \lambda^n \prod i = 1^n e^{-X_i\lambda} \\
&= \lambda^{n-1} e^{-(\sum_{i=1}^{n} X_i)\lambda}
\end{aligned}$$

As $n = 10$ and $\sum_{i=1}^{n} X_i = 12$, plugging into the formula above gives

$$\begin{aligned}
\pi(\lambda|X_1, \ldots, X_n) &\propto \lambda^{n-1} e^{-(\sum_{i=1}^{n} X_i)\lambda} \\
&= \boxed{\lambda^9 e^{-12\lambda}}.
\end{aligned}$$

- **Bayes' and MAP estimators:** First, we attempt to match the calculated expression for the posterior to a well-known distribution. Matching the posterior distribution in λ (in proportionality notation) which is

$$\pi(\lambda|X_1, \ldots, X_n) \propto \lambda^9 e^{-12\lambda}$$

to the given form for the Gamma distribution in x, which is

$$h(x) \propto x^{\alpha-1} e^{-\beta x}$$

to conclude that our posterior is a Gamma distribution with parameters $\alpha = 10$ and $\beta = 12$.

One can look up that the Gamma distribution parametrized α and β with $h(x) \propto x^{\alpha-1} e^{-\beta x}$ has mean $\frac{\alpha}{\beta}$ and mode $\frac{\alpha-1}{\beta}$. The Bayes' estimator is defined as the posterior mean, so

$$\hat{\lambda}^{\text{Bayes}} = \frac{\alpha}{\beta} = \frac{10}{12} \approx \boxed{0.833}.$$

The maximum-a-posteriori (MAP) estimator is defined as the mode of the posterior distribution, so

$$\hat{\lambda}^{\text{MAP}} = \frac{\alpha-1}{\beta} = \frac{10-1}{12} = \boxed{0.750}.$$

- **Bayesian confidence regions:** Recall from the previous item that the posterior distribution is Gamma $(10, 12)$. As $\mathbb{P}(X \geq a) = 0.95$, a is the 0.05-quantile of Gamma $(10, 12)$. Also, as $\mathbb{P}(0 < X < b) = 0.95$, b is the 0.95-quantile of Gamma $(10, 12)$. Using computational tools, such as Excel's GAMMAINV, we get that

$$a = \text{GAMMAINV}(0.05, 10, \frac{1}{12}) = \boxed{0.452},$$

and

$$b = \text{GAMMAINV}(0.95, 10, \frac{1}{12}) = \boxed{1.309}.$$

Note: There are two commonly-used parametrizations for the Gamma distribution, one that has the expression $h(x) \propto x^{\alpha-1}e^{-\beta x}$ another with expression of the form $g(x) \propto x^{\alpha-1}e^{-\frac{x}{\beta}}$. In the first form, the mean is $\frac{\alpha}{\beta}$ and the second parameter β is called the *rate* parameter, while in the second form, the mean is $\alpha\beta$ and the second parameter β is called the *scale* parameter.

When using computational tools for the Gamma distribution, it is strongly advised to perform simple sanity checks to determine which parametrization is used by the software. An easy way to check is to check the range of values when α and β are large numbers close to each other: if the *rate* interpretation for β is used, the mean will be approximately 1, while if the *scale* interpreation for β is used, the mean will be large.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(b) Bayesian Estimation and Confidence Regions: Concept Questions

4/4 points (graded)

For this section, use the answers from the previous section when referenced. We stick to the scenario where we observe $n = 10$ random variables with sum $\sum_{i=1}^{10} X_i = 12$.

Compare the values of the Bayes' estimator $\hat{\lambda}^{\text{Bayes}}$ and the maximum a-posteriori estimator $\hat{\lambda}^{\text{MAP}}$ calculated earlier. Which of the following statements correctly *describes* the relative values and provides an accurate justification for it?

- $\hat{\lambda}^{\text{Bayes}} > \hat{\lambda}^{\text{MAP}}$ because this inequality will always hold regardless of the prior or model choice.
- $\hat{\lambda}^{\text{Bayes}} < \hat{\lambda}^{\text{MAP}}$ because this inequality will always hold regardless of the prior or model choice.
- $\hat{\lambda}^{\text{Bayes}} > \hat{\lambda}^{\text{MAP}}$ because the posterior distribution is skewed to the left ✔
- $\hat{\lambda}^{\text{Bayes}} < \hat{\lambda}^{\text{MAP}}$ because the posterior distribution is skewed to the left
- $\hat{\lambda}^{\text{Bayes}} > \hat{\lambda}^{\text{MAP}}$ because the posterior distribution is skewed to the right ✔
- $\hat{\lambda}^{\text{Bayes}} < \hat{\lambda}^{\text{MAP}}$ because the posterior distribution is skewed to the right

Grading Note: Since the definition of skewed to the right or skewed to the left was not clearly defined anywhere, we will accept two answers as correct. See solution.

In a Bayesian setting, the posterior distribution can be used as a prior distribution for a succeeding Bayesian inference problem. In other words, we can write the prior $\pi_1(\lambda) = \pi(\lambda|X_1, \ldots, X_n)$. Consider the two priors: the Jeffreys prior $\pi_J(\lambda)$ earlier and the new prior $\pi_1(\lambda)$ based on the calculated posterior. Which of the two priors, if any, is/are proper?

- ○ Both $\pi_J(\lambda)$ and $\pi_1(\lambda)$
- ○ $\pi_J(\lambda)$ only
- ◉ $\pi_1(\lambda)$ only ✔
- ○ Neither $\pi_J(\lambda)$ nor $\pi_1(\lambda)$

Consider the Bayesian confidence region boundary values a and b calculated in the previous section, and now construct the interval $\mathcal{R} = (a, b)$. Which of the following statements about $\mathcal{R}$ is/are true? (Check all that apply.)

- ☐ $\mathcal{R}$ is a Bayesian confidence region of level 5%.
- ☐ $\mathcal{R}$ is the unique Bayesian confidence region of level 5%.
- ☑ $\mathcal{R}$ is a Bayesian confidence region of level 10%. ✔
- ☐ $\mathcal{R}$ is the unique Bayesian confidence region of level 10%.

✔

Is the confidence region $\mathcal{R} = (a, b)$ symmetric around the Bayesian estimator $\hat{\lambda}^{\text{Bayes}}$, as defined earlier? Why or why not? (Check all that apply.)

- ☐ Yes, because by construction, confidence intervals and hence confidence regions are symmetric around any consistent estimate of the parameter.
- ☐ Yes, because our posterior distribution is symmetric and we chose a and b such that $(-\infty, a)$ and (b, ∞) have an equal 5% probability.
- ☑ No, because our posterior distribution is not symmetric (it is either skewed to the left or skewed to the right). ✔
- ☐ No, because $\mathcal{R}$ is symmetric around the MAP, not around $\hat{\lambda}^{\text{Bayes}}$

✔

Grading note: We will give partial credit here.

Solution:

- **Bayes vs MAP estimator:** First, we note that both $\hat{\lambda}^{\text{Bayes}}$ and $\hat{\lambda}^{\text{MAP}}$ are *statistics* of the same posterior distribution $\pi(\lambda|X_1, \ldots, X_n)$. As a result, their values depend only on this distribution, which is computed from the prior and the likelihood model. From the previous question, we have computed $\hat{\lambda}^{\text{Bayes}} = 0.833$ and $\hat{\lambda}^{\text{MAP}} = 0.750$, so the appropriate comparison is $\hat{\lambda}^{\text{Bayes}} > \hat{\lambda}^{\text{MAP}}$.

  The posterior distribution $\textsf{Gamma}(10, 12)$ has a long right tail, making it skewed to the right. Hence, due to large positive values (and the absence of large negative values below 0) bringing the mean up without any effect to the mode, the mean would be larger than the mode in this posterior, explaining the observed comparison. This comparison, however, does not always hold: if the posterior is skewed to the left, then the presence of very low values (and the relative absence of very high values) will pull down the mean without affecting the mode. Thus, "the posterior distribution is skewed to the right" is an appropriate explanation for $\hat{\lambda}^{\text{Bayes}} > \hat{\lambda}^{\text{MAP}}$.

- **Proper priors:** $\pi_J(\lambda) \propto \frac{1}{\lambda}$ is an *improper prior* because the integral $\int \frac{1}{\lambda} = \ln\lambda + C$ diverges. This can also be seen by showing that the area under the curve under the graph of $\frac{1}{\lambda}$ for $\lambda > 0$ is infinite.

  On the other hand, $\pi_1(\lambda) \propto \lambda^9 e^{-12\lambda}$ is a proper prior because the area under the curve is finite as

$$\int_0^\infty \lambda^9 e^{-12\lambda} = \frac{\Gamma(b+1)}{12^{10}},$$

based on the normalized PDF of the Gamma distribution which is always well-defined.

- **Bayesian confidence region levels:** As $\mathbb{P}(X > a) = 0.95$ and $\mathbb{P}(0 < X < b) = 0.95$, with $0 < a < b$,

$$\begin{aligned}\mathbb{P}(a < X < b) &= \mathbb{P}(0 < X < b) - \mathbb{P}(0 < X < a)\\ &= 0.95 - (1 - \mathbb{P}(X > a))\\ &= 0.95 - (1 - 0.95) = 0.90 = 90\%.\end{aligned}$$

  Hence, (a, b) is a Bayesian confidence region of level 10%.

  It is, however, not the unique Bayeisan confidence region of level 90% because as long as a and b are chosen such that

$$\mathbb{P}(X > a) + \mathbb{P}(0 < X < b) = 1.90$$

  , the above computation would result to $\mathbb{P}(a < X < b) = 0.90$. For example, we could choose a and b such that $\mathbb{P}(X > a) = 0.92$ and $\mathbb{P}(0 < X < b) = 0.98$. Based on the definition of a Bayesian confidence region, $\mathcal{R}$ does not even need to be a single continuous interval constructed this way.

- **Symmetry of (a, b):** From the previous parts, we get that $a \approx 0.452$, $b \approx 1.309$, and $\hat{\lambda}^{\text{Bayes}} \approx 0.833$. $\frac{a+b}{2} \approx 0.881 \neq \hat{\lambda}^{\text{Bayes}}$, so (a, b) is not symmetric around $\hat{\lambda}^{\text{Bayes}}$. (It is also not symmetric around $\hat{\lambda}^{\text{MAP}} = 0.750$ as $\frac{a+b}{2} \neq 0.750$.)

  We don't expect the Bayesian confidence region to be symmetric around any particular estimate because the posterior distribution itself is skewed to the right. In a perfectly symmetric distribution, the distribution is symmetric around the mean, so as a and b as defined, they would be equally spaced from the mean. In an asymmetric (particularly, skewed) distribution, however, having more mass on one side of the distribution and having long tails on one side would skew the intervals towards containing more area in one side of the estimator.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(c) Frequentist Estimation and Hypothesis Testing: Large Sample

7/7 points (graded)

Now, suppose that we have $n = 100$ observations $X_1, \ldots, X_{100}$ with $\sum_{i=1}^{100} X_i = 120$. Recall $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Exp}(\lambda)$.

Compute the maximum likelihood estimate (MLE).
(Enter **numerical** answers accurate up to at least 3 decimal places.)

$\hat{\lambda}^{\text{MLE}} =$ 10/12 ✔ **Answer:** 0.833

Compute the method of moments estimate.
(Enter **numerical** answers accurate up to at least 3 decimal places.)

$\hat{\lambda}^{\text{MM}} =$ 10/12 ✔ **Answer:** 0.833

Use the plug-in method to construct a confidence interval$\mathcal{I}_{\text{plug-in}}$ for λ of asymptotic confidence level 95% centered around $\hat{\lambda}^{\text{MLE}}$. Use the variance obtained from the asymptotic variance formula for the MLE and plug in $\hat{\lambda}^{\text{MLE}}$ for λ. Enter the lower and upper bounds of (the realization of) the confidence interval below.

(Enter **numerical** answers accurate up to at least 3 decimal places.)

$\mathcal{I}_{\text{plug-in}} = [a, b]$ where

$a =$

0.67

✔ **Answer:** 0.670

$b =$

0.99667

✔ **Answer:** 0.997

Next, we decide to test the hypothesis H_0: $\lambda = 1$. Perform Wald's test using the test statistic

$$nI(\hat{\lambda}^{\mathrm{MLE}})\left(\hat{\lambda}^{\mathrm{MLE}} - \lambda_0\right)^2$$

where $I(\hat{\lambda}^{\mathrm{MLE}})$ is the Fisher information evaluated at $\hat{\lambda}^{\mathrm{MLE}}$.

Compute the p-value of Wald's test on our observations and model.

(Enter a numerical answer accurate to at least 3 decimal places.)

$p =$ 0.0455 ✔ **Answer:** 0.045

Another test we could use to test the same hypothesis is the Likelihood Ratio Test (LRT). Compute the p-value of the likelihood ratio test on our observations and model.

(Enter a numerical answer accurate to at least 3 decimal places.)

$p =$ 0.06006 ✔ **Answer:** 0.060

Suppose that we want our test to have asymptotic level $\alpha = 0.05$. Decide whether Wald's test and/or the Likelihood Ratio Test would reject the null hypothesis H_0.

- (●) Only Wald's test rejects the null hypothesis. ✔
- ( ) Only LRT rejects the null hypothesis.
- ( ) Both Wald's test and LRT reject the null hypothesis.
- ( ) Neither Wald's test or LRT reject the null hypothesis.

Solution:

- **Maximum Likelihood Estimate:** The Maximum Likelihood Estimate (MLE) for λ is the value of λ that would maximize the likelihood function $L(X_1, \ldots, X_n|\lambda)$. In a previous part, we have calculated

$$L(X_1, \ldots, X_n|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n X_i} = \lambda^{100} e^{-120\lambda}.$$

Maximizing the likelihood function is equivalent to maximizing the log likelihood, which is $\log L(X_1, \ldots X_n|\lambda) = 100 \log \lambda - 120\lambda$. This goes to $-\infty$ as $\lambda \to 0$ and $\lambda \to \infty$, and is continuous, so the maximum occurs when we set $\frac{d}{d\lambda} \log L(X_1, \ldots, X_n|\lambda) = 0$. This is equivalent to $\frac{100}{\lambda} - 120 = 0$, or $\lambda = \frac{100}{120} \approx \boxed{0.833}$.

- **Method of Moments Estimate**: We are finding $\hat{\lambda}$ such that

$$\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n X_i.$$

MM的估计是这样算的：找出某一moment和参数的关系，再用样本均数替代期望

$\mathbb{E}[X] = \frac{1}{\lambda}$, and $\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{100} 120$, so we get $\frac{1}{\hat{\lambda}} = \frac{120}{100}$ and thus $\hat{\lambda} = \frac{100}{120} \approx \boxed{0.833}$.

- **MLE Plug-in Confidence Interval:** The asymptotic normality result for $\hat{\lambda}^{\text{MLE}}$ states that

$$\sqrt{n}\,(\hat{\lambda}^{\text{MLE}} - \lambda) \overset{n\to\infty}{\to} \mathrm{N}\left(0, I(\lambda)^{-1}\right),$$

where $I(\lambda)$ denotes the Fisher information.

We have computed $I(\lambda) = \frac{1}{\lambda^2}$, so the asymptotic variance is $I(\lambda)^{-1} = \lambda^2$. Now, we could estimate the variance of $\hat{\lambda}^{\text{MLE}}$ for $n = 100$ by plug-in. Assuming that the asymptotic variance result holds for $n = 100$, we get the variance of $\frac{\lambda}{100}$ by moving $\sqrt{n}$ to the right hand side. Then, we could estimate λ by $\hat{\lambda}$ to get an estimated variance (for $\hat{\lambda}^{\text{MLE}}$) of

$$\frac{\hat{\lambda}^2}{100} \approx \frac{0.8333^2}{100} = 0.006944.$$

We then get an estimated standard error of $\hat{\text{SE}} = \sqrt{0.006944} \approx 0.08333.$

From this, we can calculate the endpoints of our plug-in confidence interval to be

$$a = \hat{\lambda}^{\text{MLE}} - 1.96\hat{\text{SE}} = 0.8333 - 1.96\,(0.8333) \approx \boxed{0.670}$$

and

$$b = \hat{\lambda}^{\text{MLE}} + 1.96\hat{\text{SE}} = 0.8333 + 1.96\,(0.8333) \approx \boxed{0.997}.$$

- **Wald's Test:** Recall that $\hat{\lambda}^{\text{MLE}} = \frac{5}{6} \approx 0.8333$, and $I(\lambda) = \frac{1}{\lambda^2}$ so $I(\hat{\lambda}^{\text{MLE}}) = 1.44$. Hence the Wald test statistic is

$$nI(\hat{\lambda}^{\text{MLE}})\,(I(\hat{\lambda}^{\text{MLE}}) - \lambda_0)^2 = 100\,(1.44)\,(\frac{5}{6} - 1)^2 = 144(\frac{-1}{6})^2 = 4.$$

The asymptotic distribution of the Wald's test statistic under the null hypothesis is χ^2_1, and using computational tools gives that a test statistic value of 4 is the 0.9545-quantile of the distribution. Hence, the p-value of the test is $1 - 0.9545 = \boxed{0.0455}$. As $0.0455 < 0.05$, we reject H_0 at the $\alpha = 0.05$ level through Wald's test.

- **Likelihood Ratio Test (LRT):** We compute both the maximum log likelihoods in both the unrestricted and restricted cases, where the restriction is $\lambda = 1$. The log likelihood is $n\log\lambda - \lambda\sum_{i=1}^{n} X_i$. In the unrestricted case, the MLE occurs when $\lambda = \frac{5}{6}$, so with $n = 100$ and $\sum_{i=1}^{n} X_i = 120$, the maximum log likelihood is

H0

$$l_n(\hat{\lambda}) = 100\log\frac{5}{6} - \frac{5}{6}120 \approx -118.2321$$

. In the restricted case, we need to have $\lambda = 1$, so the log likelihood is

$$l_n(\hat{\lambda}^c) = 100\log 1 - 120 = -120.$$

The LRT test statistic is $T_n = 2\,(l_n(\hat{\lambda}) - l_n(\hat{\lambda}^c))$, so taking the computed values gives $T_n = 2\,(-118.2321 - (-120)) = 3.536$. The null hypothesis tests only one variable, so the distribution of T_n under H_0 is χ^2_1. Using computational tools gives that a test statistic value of 3.536 is the 0.9399-quantile of the χ^2_1 distribution. Hence, the p value of the test is $1 - 0.9399 = \boxed{0.0601}$. As $0.0601 > 0.05$, we fail to reject H_0 at the $\alpha = 0.05$ level through the LRT.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(d) Small Sample Issues

2/2 points (graded)

In this section, we consider the case where we have a small sample, but still wish to either estimate λ, provide a range for it, or perform a hypothesis test. Here, we consider the scenario in the first section, where we have $n = 10$ observations $X_1, \ldots, X_n$ such that

$$\sum_{i=1}^{10} X_i = 12.$$

Select all true statements about **small-sample** estimation and inference below.

- [x] The following estimators are all well-defined in the small sample case: MLE, Method of Moments, Bayes' estimator, MAP estimator. ✔

- [ ] A confidence interval of non-asymptotic confidence level 5% can be constructed using the plug-in method and the asymptotic variance for $\hat{\theta}^{\text{MLE}}$, as done above in part (c).

✔

Grading note: Partial credit is given.

Using Student's t-test for testing the hypothesis $\lambda = 1$ will provide a reliable, non-asymptotic result for $n = 10$ observations with an exponential likelihood.

- ( ) True, as Student's t-test is meant handle hypothesis testing with a small sample size.
- ( ) True, as the mean of multiple i.i.d. exponential random variables is necessarily a Gaussian distribution.
- ( ) False, because Student's t-test can only test whether a parameter is zero.
- (•) False, because one of Student's t-test assumptions is that the observations are drawn from a Gaussian distribution. ✔

Solution:

- **Estimators well-defined:** All four estimators are well-defined, even for the small sample case. We have seen earlier that for an exponential model, both the MLE and the Method of Moments estimator are both equivalent to $\frac{n}{\sum_{i=1}^{n} X_i}$. The other two Bayesian estimators are based on the posterior distribution, which can be calculated even in the small-sample case.

- **Non-asymptotic confidence level:** The statement is false as the proposed method only generates *asymptotic* confidence intervals for $\hat{\lambda}^{\text{MLE}}$, not non-asymptotic confidence intervals. An asymptotic confidence interval for a very small sample is not reliable as the plug-in method uses the asymptotic normality result for the MLE, which is only a good approximation for large samples.

- **Student's t-test:** The statement is false, because the Student's t-test assumes that the model is Gaussian; this assumption is needed to produce the t-distribution under the null hypothesis. It is true, however, that the Student's t-test is meant to handle hypothesis testing with a small sample size and produces a non-asymptotic result, though this property relies on the Gaussian distribution assumption. Lastly, the t-test may be used to test whether the parameter is any particular real number, not just whether it is zero.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(e) Goodness-of-fit Tests

1/3 points (graded)

Suppose a statistician wishes to test whether a large number of observations X_i follows an exponential distribution with parameter $\lambda = 1$. He wishes to test this hypothesis exactly, and intends that if the observations follow an exponential distribution with a different parameter, the test should reject the null hypothesis given sufficiently many observations. In addition, he wants to have a numeric statistic that he could report and does not want the procedure to involve rounding off observation numbers into bins. Which of the following goodness-of-fit tests would be the most appropriate for this purpose?

- Chi-squared Test ✖
- Kolmogorov-Smirnov Test ✔
- Kolmogorov-Liliefors Test
- Quantile-quantile plots

Another statistician intends to assign an undergraduate to perform the chi-squared test and plot QQ-plots to test the hypothesis $H_0 : \lambda = 1$. Suppose that he does not get to see the actual data; he has only the test result as well as the plot. He thinks that the data may actually be generated from a geometric distribution with parameter $p = 1 - e^{-1}$, and is worried that the tests he had in mind cannot distinguish between the distributions $\text{Exp}(1)$ and $\text{Geom}(1 - e^{-1})$, where $\text{Geom}(1 - e^{-1})$ is defined by the pdf $p(k) = (1-p)^k p$, for $k = 0, 1, \ldots$.

Which of the followng is true if he uses the chi-squared test in this situation? (Choose all that apply.)

- [x] If he uses the chi-squared test with bins $[0, 1)$, $[1, 2)$, ..., $[N-1, N)$, $[N, \infty)$ for some N, then the two distributions are indistinguishable. ✔
- [ ] If he uses the chi-squared test with bins that all contain at least one positive integer, then the two distributions are indistinguishable.

✔

Grading note: The definition of the geometric distribution was ambiguous without the pdf. However, since the second choice is false, and this type of problem requires at least 1 check mark, we consider this to be an acceptable problem. Partial credit is given.

Which of the followng is true if he uses the QQ-plot in this situation?

看错了

- It will not be possible to tell the distributions apart using a QQ-plot given a sufficiently large number of samples. ✖
- It will be possible to tell the distributions apart using a QQ-plot given a sufficiently large number of samples. ✔

Correction Note: June 2 An earlier version of the problem statement did not include the pdf of the geometric distribution.

Solution:

- **Choosing a Goodness-of-fit Test:** The Kolmogorov-Smirnov test is the most appropriate as it tests for a specific model with a fixed parameter, unlike the Kolmogorov-Liliefors test which allows the parameter to be flexible (by plugging in the MLE).

  While the Chi-squared test and QQ-plots may be useful to perform such a test, they do not follow the specificiations in the problem statement. The Chi-squared goodness-of-fit test for a continuous distribution requires rounding off observation numbers into bins (as it's applied to discrete events). On the other hand, QQ-plots are a qualitative comparison of two distributions and do not produce any numeric statistic.

- **Chi-squared test distinguishability:** The chi-squared test is unable to distinguish between two distributions, under a certain binning system, if for each bin, there is an equal probability for a random variable drawn from either distribution to fall in the bin.

  The first statement is true. If the distribution is $\text{Exp}(1)$, the probability of falling in the interval $[i, i+1)$ is $e^{-i} - e^{-i-1} = e^{-i}(1 - e^{-1})$. (This is because the CDF of $\text{Exp}(1)$ is e^{-x}.) On the other hand, if the distribution is $\text{Geom}(1 - e^{-1})$, the probability of falling in the interval $[i, i+1)$ is the probability of the random variable being i, which is $(1 - (1 - e^{-1}))^i (1 - e^{-1}) = e^{-i}(1 - e^{-1})$, which is the same as in the $\text{Exp}(1)$ model.

  The second statement is false. Consider the bin $[0, 1.5)$, which indeed contains at least one positive integer. If the distribution is $\text{Exp}(1)$, the probability of falling in $[0, 1.5)$ is $1 - e^{-1.5} \approx 0.777$. If the distribution is $\text{Geom}(1 - e^{-1})$, it's the probability that the random variable is either 0 or 1, so the total probability is $(1 - (1 - e^{-1}))^0 (1 - e^{-1}) + (1 - (1 - e^{-1}))^1 (1 - e^{-1}) \approx 0.865$. The probabilities are different, so a Chi-squared test would be able to distinguish between the distributions given a sufficient number of samples.

- **QQ-plot distinguishability:** It is possible to tell the distributions apart given a sufficient number of samples. The distribution Geom $(1 - e^{-1})$ is discrete, so on a QQ-plot against Exp (1), if the distribution is actually from Geom $(1 - e^{-1})$, the set of values attained (and hence the quantiles in Exp (1) will be from a discrete set. Thus the points will definitely not lie in a diagonal line as more samples are obtained; it will resemble a set of disjoint vertical sets going up from left to right.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Show Discussion

Topic: Final exam:Final Exam / Problem 3: Review through an Exponential Model