M347 TMA02 R2698663

Louis Botterill

07-01-2018

Question 1

1.a

$x_1, x_2, ..., \ x_n$ are independent observations taken from a Pareto distribution with pdf

$$f(x|\beta) = \frac{\beta}{x^{\beta+1}} \text{ on } x > 1; \text{ where } \beta > 0$$

We can show the log-likelihood is as required as follows

Definition: Log-likelihood is defined as:

$$\boxed{l(\theta) = log\ L(\theta) = \sum_{i=1}^{n}\ log\ f(x_i|\theta)}$$

So we have

$$l(\beta) = \sum_{i=1}^{n}\ log\ \frac{\beta}{x_i^{\beta+1}}$$
$$l(\beta) = n\ log\ \beta + \sum_{i=1}^{n}\ log\ \frac{1}{x_i^{\beta+1}}$$
$$l(\beta) = n\ log\ \beta - \sum_{i=1}^{n}\ log\ x_i^{\beta+1}$$
$$l(\beta) = n\ log\ \beta - (\beta+1)\sum_{i=1}^{n}\ log\ x_i$$

1.b

Find $l'(\beta)$ and hence the candidate MLE

$$l'(\beta) = \frac{d}{d\beta} l(\beta) = d\ [log\ \beta - (\beta+1)\sum_{i=1}^{n}\ log\ x_i]/d\beta$$
$$l'(\beta) = \frac{n}{\beta} - \sum_{i=1}^{n}\ log\ x_i$$

Solving for the stationary point where $l'(\theta) = 0$

$$\frac{n}{\beta} - \sum_{i=1}^{n}\ log\ x_i = 0$$

$$\frac{n}{\beta} = \sum_{i=1}^{n} \ log\ x_i$$

So the candidate MLE is

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ log\ x_i}$$

(At this point $\hat{\beta}$ it is only a candidate MLE of β since we've not proved $\hat{\beta}$ is a maximum, it could be a saddle or minimum.)

1.c

We can confirm that $\hat{\beta}$ is indeed the MLE of β by showing that the stationary point of the function $l'(\beta)$ is a maximum. This occurs when the second derivative $l''(\beta) < 0$

$$\begin{aligned} l'(\beta) &= n\beta^{-1} - \sum_{i=1}^{n} \ log\ x_i \\ l''(\beta) &= -n\beta^{-2} = -\frac{n}{\beta^2} \end{aligned}$$

Since $n > 0$ (and $\beta > 0$), then $l''(\beta) < 0$, for all n, β

Hence,

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ log\ x_i}$$

is an MLE of β

1.d

Explain why $\hat{\beta} > 0$, as one would hope given that $\beta > 0$

We have found that the MLE $\hat{\beta}$ is defined as

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ log\ x_i}$$

where $n > 0$, positive ($n = 0$ would be a degenerate case) and also $x > 1$, positive (from the definition of the distribution)

Definition: By definition of logarithms

$$\boxed{ln(x > 1) > 0}$$

and so the result of positive n times 1/sum of positive values is always positive, as required

Question 2

x_1, x_2 ... x_n are independent observations taken from a Pareto distribution with pdf

$$f(x|\beta) = \frac{\beta}{x^{\beta+1}} \text{ ; on } x > 1 \text{ where } \beta > 0$$

$$l(\beta) = n \; log \; \beta - (\beta+1)\sum_{i=1}^{n} \; log \; x_i$$

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \; log \; x_i}$$

Given

$$E\big\{ - l''(\beta)\big\} = \frac{n}{\beta^2}$$

and

The hypothesis test is: $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$.

So $\beta_0 = 1$ and $\beta_1 \neq 1$

2.a

Obtain formulae for both Wald statistics, W_1 and W_2, in terms of β (and n).

Definition: The Wald W_1 and W_2 statistics on parameter θ

$$\boxed{\begin{aligned} W_1 &= (\theta_0 - \hat{\theta})^2 E\big\{ - l''(\theta)\big\}_{\hat{\theta}} \\ W_2 &= (\theta_0 - \hat{\theta})^2 E\big\{ - l''(\theta)\big\}_{\theta_0} \end{aligned}}$$

So we have

$$\begin{aligned} W_1 &= (1 - \hat{\beta})^2 \frac{n}{\hat{\beta}^2} \\ W_2 &= (1 - \hat{\beta})^2 \frac{n}{\beta_0^2} = n(1 - \hat{\beta}^2) \end{aligned}$$

Using $\beta_0 = 1$

2.b

Show that the score statistic, S, can be written

$$S = (1-\hat{\beta})^2 \ \frac{n}{\hat{\beta}^2}$$

Definition: The Score test statistic is defined as

$$\boxed{S = \frac{\{l'(\theta_0)\}^2}{E\{-l''(\theta)\}_{|\theta_0}}}$$

From question part 1.b

$$l'(\beta) = \frac{n}{\beta} - \sum_{i=1}^{n} \ log \ x_i$$

$$l'(\beta_0 = 1) = \frac{n}{\beta_0} - \sum_{i=1}^{n} \ log \ x_i = n - \sum_{i=1}^{n} \ log \ x_i$$

$$E\{-l''(\beta)\}|_{\beta_0=1} = \frac{n}{\beta_0^2} = n$$

and

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ log \ x_i}$$

$$\sum_{i=1}^{n} \ log \ x_i = \frac{n}{\hat{\beta}}$$

$$l'(\beta) = n - \frac{n}{\hat{\beta}}$$

$$l'(\beta) = n(1 - \frac{1}{\hat{\beta}})$$

$$(l'(\beta))^2 = (n - \frac{n}{\hat{\beta}})^2 = n^2(1 - \frac{1}{\hat{\beta}})^2$$

$$S = n^2 \frac{(1-\frac{1}{\hat{\beta}})^2}{n} = n(1 - \frac{1}{\hat{\beta}})^2$$

Which we can rewrite using

$$(1 - \frac{1}{\beta})^2 = (1 + \frac{1}{\beta^2} - \frac{2}{\hat{\beta}}) = \frac{1}{\hat{\beta}^2}(\hat{\beta}^2 + 1 - 2\hat{\beta}) = \frac{1}{\hat{\beta}^2}(1 - \hat{\beta})^2$$

Therefore

$$S = (1 - \hat{\beta})^2 \frac{n}{\hat{\beta}^2}$$

As required.

2.c

The common procedure to evaluate the tests and obtain the appropriate p-value would be:

1 - Evaluate the Wald W_1, W_2 or Score S test statistic value, r say, using the previously derived test statistic formulae and observed sample data

2 - Calculate the p-value from r using the appropriate asymptotic null distribution, in this case this is $\chi^2(d=1)$

- The p-value is given by $p = P(Y \geq r)$, where $Y \sim \chi^2(1)$ and r is the observed value of $2log(LR)$ which in our case is approximated by our test statistics r
- The p-value $P(X^2(d=1) \geq r)$ would be $1 - P(\chi^2(d=1) < r)$ or 1-F(r) where F is the cdf of $\chi^2(d=1)$, the asymptotic null distribution

For example, if r was 3.84 we would find the p-value $= \chi^2_{d=1}(x \geq 3.84) = 1 - \chi^2_{d=1}(x < 3.84) = 1 - 0.95 = 0.05$ (this is the probability of getting a result at least this extreme, if the null hypothesis is true)

3 - Decide whether to accept or reject the null hypothesis by comparing the p-value against a probability threshold α:

- Lower p-values mean the test statistic value is more unlikely if the null hypothesis is true, so we have more evidence to reject the null hypothesis H_0, so we might reject the null hypothesis as false, more readily.
- Higher p-values mean the test statistic value is more likely if the null hypothesis is true, so we have less evidence to reject the null hypothesis H_0, we might accept the null hypothesis as true, more readily.

The justification for the using the $\chi^2(d=1)$ is due to the fact that the asymptotic null distribution of the likelihood ratio (LR) test statistic approaches $\chi^2(d)$, where d is the degrees of freedom difference between the null and alternative hypothesis.

i.e. $2log(LR) \approx \chi^2(d)$, if the null hypothesis H_0 is true

(This approximation improves as the sample size n increases)

In our example, we have H_0 with zero free parameters so $d_0 = 0$ and H_1 with 1 free parameter β_1 so $d_1 = 1$. Therefore d = 1 in our case, since there is 1 degree of freedom difference between the null H_0 and alternative H_1 hypothesis. $d = d_1 - d_0 = 1 - 0 = 1$

So the $\chi^2(d=1)$ distribution is the correct distribution to use to compare with the test statistic of the observed sample data.

If however, we started out with a fixed-level test size in mind, a similar approach using the appropriate quantile of the same $\chi^2(d=1)$ distribution would be:

Compare the observed value of the test statistic r with the $\chi^2(1)$ distribution. Similarly, reject the null hypothesis H_0 if the observed value of the test statistic r is too large. For example, a fixed-level test of size $\alpha = 0.05$ will reject H_0 if the test statistic value r is greater than the $1 - 0.05 = 0.95^{th}$-quantile of the $\chi^2(1)$ distribution, which has a value of 3.84.

Further detail (also for my revision/understanding purposes!)

1 - Calculate the test statistic T from the observed values.

- If the null hypothesis is true, these test statistics T approximate 2 log(LR) from the observed data. They are equivalent asymptotically to the likelihood ratio test, as the sample size grows large its asymptotic null distribution is $\chi^2(d)$

2 - Use the appropriate $\chi^2(d)$ and find the critical value k to be compared with the test statistic.

- $k = \chi^2_{1-\alpha}(d)$, where $\chi^2_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the $\chi^2(d)$ distribution
- An example would be for a test size $p = 0.05$ we would find the $1 - p = 0.95$ quantile of the $\chi^2(d)$ distribution
- Where the p-value is the probability of observing data at least as extreme as those obtained, if the null hypothesis were true. Hence, p-value $= P(\chi(d)^2 \geq 2log(LR_0)|H_0 = true)$ where $log(LR_0)$ is the observed value of the log-likelihood ratio test statistic.

3 - Compare the test statistic value r with the critical value k from the selected quantile of $\chi^2(d = 1)$

- If the test statistic is too large, reject the null hypothesis H_0
- If the test statistic is not too large, we accept the null hypothesis H_0

Definition: The p-value is defined as

$$\boxed{p = P(log(LR) \geq log(LR_0)|H_0\ true)}$$

The p-value is given by $p = P(Y \geq r)$, where $Y \approx \chi^2(d)$ and r is the observed value of 2 log(LR) (in our case the value of the test statistic approximation thereof).

Question 3

3.a

Given

$$X_n = X + \frac{Z_n}{n^{1/2}}$$

and

$$E(Z_n^2) = cn^{\alpha} \text{ ; where } c > 0, \alpha \in \mathbb{R}$$

Using Chebyshev's inequality, show that

$$P(|X_n - X| > \epsilon) \leq \frac{c}{\epsilon^2 n^{1-\alpha}}$$

Definition: Chebyshev's inequality is defined as:

$$\boxed{P(|X| \geq a) \leq \frac{E(X^2)}{a^2}}$$

So we have

$$\begin{aligned}
P(|X_n - X| > \epsilon) &= P(|\frac{Z_n}{n^{1/2}}| > \epsilon) \\
&= P(|Z_n| > \epsilon n^{1/2}) \\
&\leq P(|Z_n| \geq \epsilon n^{1/2}) \\
&\leq \frac{E(Z_n^2)}{(\epsilon n^{1/2})^2} \text{ (using Chebyshev's inequality)} \\
&= \frac{cn^{\alpha}}{(\epsilon n^{1/2})^2} = \frac{cn^{\alpha}}{\epsilon^2 n} \\
&= \frac{cn^{\alpha} n^{-1}}{\epsilon^2} = \frac{cn^{\alpha-1}}{\epsilon^2} \\
P(|X_n - X| > \epsilon) &= \frac{c}{\epsilon^2 n^{1-\alpha}}
\end{aligned}$$

Therefore we have shown that

$$P(|X_n - X| > \epsilon) \leq \frac{c}{\epsilon^2 n^{1-\alpha}}$$

As required.

3.b

For what values of α does the argument in part (a) prove that X_n converges in probability to X ?

We have

$$P(|X_n - X| > \epsilon) \leq \frac{c}{\epsilon^2 n^{1-\alpha}}$$

We can conclude X_n tends to X by probability convergence (as $n \to \infty$) when

$$P(|X_n - X| > \epsilon) \to 0$$

$$\frac{c}{\epsilon^2 n^{1-\alpha}} \to 0$$

$$\frac{c}{\epsilon^2 n^{1-\alpha}} = \frac{c}{\epsilon^2} n^{\alpha-1}$$

As $n \to \infty$ we require $n^{\alpha-1} \to 0$

$n^k \to 0$ as $n \to \infty$ for $k < 0$ so for probability convergence we require

$\alpha - 1 < 0$, therefore when $\alpha < 1$ we have $X_n \to X$ by probability convergence.

3.c

For the values of α identified in part (b), what other mode of convergence of X_n to X is assured (without any further calculations)?

We can use the rules of implication for convergence

Definition:

$$\{X_n \xrightarrow{ms} X\} \implies \{X_n \xrightarrow{p} X\} \implies \{X_n \xrightarrow{D} X\}$$

We have shown probability convergence in part 3.b

$$\{X_n \xrightarrow{p} X\} \implies \{X_n \xrightarrow{D} X\}$$

This implies also distribution convergence by the above definition of the implications of convergence.