Probability and Statistics Assessed Coursework 6

Solution 1

To verify that $f_\theta(x)$ is a valid probability density function, we need to show that for all values of x, $f_\theta(x) \geq 0$ and $\int_{-\infty}^{\infty} f_\theta(x)dx = 1$, under the condition that $\theta \in (1, \infty)$

The value of $f_\theta(x)$ is indeed always greater or equal to 0. Since $\theta > 1$, we know that $\theta - 1 > 0$. Since $x^{-\theta}$ is always positive and $f_\theta(x) = 0$ when $x < 1$, we have proved that $f_\theta(x) \geq 0$.

We now show that $\int_{-\infty}^{\infty} f_\theta(x)dx = 1$

$$\begin{aligned}
\int_{-\infty}^{\infty} f_\theta(x)dx &= \int_{-\infty}^{1} f_\theta(x)dx + \int_{1}^{\infty} f_\theta(x)dx \\
&= 0 + \int_{1}^{\infty} x^{-\theta}(\theta - 1)dx \\
&= (-x^{1-\theta})|_1^\infty \\
&= 0 - (-x^0) \\
&= 1
\end{aligned}$$

We now use the maximum likelihood estimate $\hat{\theta}$ to get the fittest value of θ. We first find the log-likelihood function and then take its derivative

$$\begin{aligned}
l(\theta|\bar{x}) &= \sum_{i=1}^{n} \ln(f_{\theta|X}(x_i)) \\
&= \sum_{i=1}^{n} \ln(x_i^{-\theta}(\theta - 1)) \\
&= \sum_{i=1}^{n} \ln(x_i^{-\theta}) + \sum_{i=1}^{n} \ln(\theta - 1) \\
&= -\theta \sum_{i=1}^{n} \ln(x_i) + n\ln(\theta - 1) \\
l'(\theta|\bar{x}) &= -\sum_{i=1}^{n} \ln(x_i) + \frac{n}{\theta - 1}
\end{aligned}$$

We then let the derivative equalt to 0 in order to solve for $\hat{\theta}$

$$\begin{aligned}
-\sum_{i=1}^{n} \ln(x_i) + \frac{n}{\hat{\theta} - 1} &= 0 \\
\frac{n}{\hat{\theta} - 1} &= \sum_{i=1}^{n} \ln(x_i) \\
\hat{\theta} &= \frac{n}{\sum_{i=1}^{n} \ln(x_i)} + 1
\end{aligned}$$

Solution 3

a) We now perform a maximum likelihood estimate $\hat{p}$ for the parameter p. We first find the log-likelihood function and then take its derivative

$$
\begin{aligned}
l(p|\bar{x}) &= \sum_{i=1}^{1469} \ln(p_{p|X}(x)) \\
&= \sum_{i=1}^{1469} \ln(p(1-p)^{x_i-1}) \\
&= \sum_{i=1}^{1469} \ln(p) + \sum_{i=1}^{1469} \ln((1-p)^{x_i-1}) \\
&= 1469\ln(p) + \ln(1-p)\sum_{i=1}^{1469}(x_i-1) \\
&= 1469\ln(p) + \ln(1-p)\sum_{i=1}^{1469} x_i - 1469\ln(1-p) \\
l'(p|\bar{x}) &= \frac{1469}{p} - \frac{\sum_{i=1}^{1469} x_i}{1-p} + \frac{1469}{1-p}
\end{aligned}
$$

We then let the derivative equal to 0 so that we can obtain the value of $\hat{p}$

$$
\begin{aligned}
\frac{1469}{\hat{p}} - \frac{\sum_{i=1}^{1469} x_i}{1-\hat{p}} + \frac{1469}{1-\hat{p}} &= 0 \\
1469(1-\hat{p}) + 1469\hat{p} &= \hat{p}\sum_{i=1}^{1469} x_i \\
1469 - 1469\hat{p} + 1469\hat{p} &= \hat{p}\sum_{i=1}^{1469} x_i \\
\hat{p} &= \frac{1469}{\sum_{i=1}^{1469} x_i} = \frac{1469}{2282} = 0.644
\end{aligned}
$$

We then check the second derivative to see if the value of $\hat{p}$ is indeed the maximum

$$
l''(p|\bar{x}) = -\frac{1469}{p^2} - \frac{\sum_{i=1}^{1469} x_i}{(1-p)^2} + \frac{1469}{(1-p)^2} = -\frac{1469}{p^2} - \frac{813}{(1-p)^2}
$$

where both terms in the equation are negative, which means that $l''(p|\bar{x}) < 0$ and $\hat{p}$ is indeed the maximum value.

b) A goodness of fit test, particularly a chi-square analysis, can be carried out to see if the underlying population distritbution is geometric. We can first construct an observed random variabel $O = (O_1, O_2, \ldots, O_n)$ where the value of O_i corresponds to the number of occurance of the value x_i of random variable X, the number of occupants in private cars on a certain road. Then we can use the chi-square statistic

$$
\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}
$$

We can then populate the table using the geometric distribution as the expected value for each count, and then calculate the value of χ^2. Then we can use the chi-square statistic table with degree of freedom 4 and two-tailed confidence level 0.99 to see the score we can obtain. If χ^2 is greater than the score, then we will reject the null hypothesis that the underlying distribution is geometric.