Solution 4: Statistical inference (I)

Siyue Yang

05/10/2022

Part 1: Probability distributions

1. A contestant on a game show needs to answer 10 questions correctly to win the jackpot. However, if they get 4 incorrect answers, they are kicked off the show. Suppose one contestant consistently has a 80% chance of correctly responding to any question.
   (a) What is the probability distribution?
   (b) What is the probability that she will correctly answer 10 questions before 4 incorrect responses?
   (c) Write out the R code to calculate (b).
2. A small town's police department issues 5 speeding tickets per month on average.
   (a) Using a Poisson random variable, what is the likelihood that the police department issues 3 or fewer tickets in one month?
   (b) What is the probability that 10 days or fewer elapse between two tickets being issued?
   (c) Write out the R code to calculate (a), (b).

Solution

1. Negative Binomial distribution.

Letting Y represent the number of incorrect responses, and setting $r = 10$, we want

$$\begin{aligned}
P(Y < 4) =& P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) \\
=& \left(\begin{array}{c} 9 \\ 9 \end{array} \right) (1 - 0.8)^0 (0.8)^{10} + \left(\begin{array}{c} 10 \\ 9 \end{array} \right) (1 - 0.8)^1 (0.8)^{10} \\
&+ \left(\begin{array}{c} 11 \\ 9 \end{array} \right) (1 - 0.8)^2 (0.8)^{10} + \left(\begin{array}{c} 12 \\ 9 \end{array} \right) (1 - 0.8)^3 (0.8)^{10} \\
=& 0.97
\end{aligned}$$

```r
sum(dnbinom(0:3, size = 10, prob = .8))
```

```
## [1] 0.7473243
```

2. First, we note that here $P(Y \leq 3) = P(Y = 0) + P(Y = 1) + \cdots + P(Y = 3)$. Applying the probability mass function for a Poisson distribution with $\lambda = 5$, we find that

$$\begin{aligned}
P(Y \leq 3) &= P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) \\
&= \frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!} + \frac{e^{-5}5^3}{3!} \\
&= 0.27.
\end{aligned}$$

```r
sum(dpois(0:3, lambda = 5))    # or use ppois(3, 5)
```

```
## [1] 0.2650259
```

We know the town's police issue 5 tickets per month. For simplicity's sake, assume each month has 30 days. Then, the town issues $\frac{1}{6}$ tickets per day. That is $\lambda = \frac{1}{6}$, and the average wait time between tickets is $\frac{1}{1/6} = 6$ days. Therefore,

$$P(Y < 10) = \int_0^{10} \frac{1}{6} e^{-\frac{1}{6}y} dy = 0.81$$

```r
pexp(10, rate = 1/6)
```

```
## [1] 0.8111244
```

Part 2: Statistical inference

1. (AoS 6.6.2) Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$ and let $\hat{\theta} = \max\{X_1, \ldots, X_n\}$. Find the bias, se and MSE of this estimator.
2. (AoS 6.6.3) Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$ and let $\hat{\theta} = 2\bar{X}_n$. Find the bias, se and MSE of this estimator.
3. Let $X_1, \ldots, X_n \sim \text{Uniform}(0, 1)$. Let $Y_n = \bar{X}_n^2$. Find the limiting distribution of Y_n. (Hint: CLT)

Solution

1. The CDF G of $\widehat{\theta}$ is

$$\begin{aligned} G(\widehat{\theta}) &= \mathbb{P}(\widehat{\Theta} \leq \widehat{\theta}) \\ &= \mathbb{P}\left(\max\{X_1, \ldots, X_n\} \leq \widehat{\theta}\right) \\ &= \prod_{i=1}^{n} \mathbb{P}\left(X_i \leq \widehat{\theta}\right) \\ &= F_\theta(\widehat{\theta})^n \\ &= \left(\frac{\widehat{\theta}}{\theta}\right)^n \end{aligned}$$

The density is therefore

$$g(\widehat{\theta}) = \left(\frac{n}{\theta}\right)\left(\frac{\widehat{\theta}}{\theta}\right)^{n-1}$$

Thus,

$$\mathbb{E}_\theta(\widehat{\theta}) = \int_0^\theta \widehat{\theta} g(\widehat{\theta}) d\widehat{\theta} = \frac{n\theta}{n+1}$$

and

$$\text{bias } = \frac{n\theta}{n+1} - \theta = -\frac{\theta}{n+1}.$$

Also,

$$\mathbb{E}_\theta\left(\widehat{\theta}^2\right) = \int_0^\theta \widehat{\theta}^2 g(\widehat{\theta}) d\widehat{\theta} = \frac{n\theta^2}{n+2}$$

and so

$$\mathbb{V}_\theta(\widehat{\theta}) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+2)(n+1)^2}$$

The mse is

$$\text{bias}^2 + \mathbb{V} = \left(\frac{\theta}{n+1}\right)^2 + \frac{n\theta^2}{(n+2)(n+1)^2} = \frac{2\theta^2}{(n+1)(n+2)}$$

2. Recall that $\mathbb{E}\,(X_i) = \theta/2, \mathbb{V}\,(X_i) = \theta^2/12$. So

$$\mathbb{E}_\theta(2\bar{X}) = 2\mathbb{E}_\theta(\bar{X}) = 2\frac{\theta}{2} = \theta$$

and hence bias $= 0$. Now

$$\mathbb{V}_\theta(2\bar{X}) = 4\mathbb{V}_\theta(\bar{X}) = \frac{4\sigma^2}{n} = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}.$$

Since this estimator is unbiased,

$$\text{mse} = \mathbb{V}_\theta(\widehat{\theta}) = \frac{\theta^2}{3n}.$$

3. $\mu = \mathbb{E}\,(X_i) = 1/2$ and $\sigma^2 = \mathbb{V}\,(X_i) = 1/12$. By the CLT,

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} = \sqrt{12n}\left(\bar{X} - \frac{1}{2}\right) \rightsquigarrow N(0,1).$$

Now $Y = g(\bar{X})$ where $g(s) = s^2$. And $g'(s) = 2s$ and $g'(\mu) = g'(1/2) = 2(1/2) = 1$. From the delta method,

$$\frac{\sqrt{n}(Y - g(\mu))}{|g'(\mu)|\,\sigma} = \sqrt{12n}\left(\bar{X} - \frac{1}{4}\right) \rightsquigarrow N(0,1)$$