Problem set 3: Summarizing distributions

Due October 21, 2024, at 10am

(Your name here)

Note: Start with the file ps3_2024_summarizing_distributions.qmd (available from the github repository at https://github.com/UChicago-pol-methods/IntroQSS-F24/tree/main/assignments). Modify that file to include your answers. Make sure you can "render" the file (e.g. in RStudio by clicking on the Render button). Submit both the qmd file and the PDF via Canvas.

Question 1: Expected value

Consider the random variable X characterized by the following PMF.

\overline{x}	f(x)
0	.3
1	.5
2	.15
3	.05

(1a) Compute E[X]. Show your work.

Answer:

(1b) Write an R function to compute the expectation of any discrete random variable. The arguments to your function should include the values the random variable can take on (x) and the probability it takes on each value (probs). Use your function to confirm your answer from question (1a).

Answer:

your code here

(1c) Compute the MSE (E $[(X-c)^2]$) for c=0 and c=1. Show your work.

Answer:

(1d) Write a function to compute the MSE for any discrete random variable at a value c. The arguments to your function should include the values the random variable can take on (x), the probability it takes on each value (probs), and the value c being considered. Use your function to confirm your answers from (1c).

Answer:

```
# your code here
```

(1e) Create a vector cs that contains numbers in the sequence 0.5, 0.55, 0.6, 0.65, 0.7, . . . , 2. Using a for-loop and your function from (1d), compute the MSE for the random variable X whose PMF was given above at each value of cs and store the result in a vector called mses.

Answer:

```
# your code here
```

(1f) Using the plot() command, make a scatterplot showing the MSE of X (vertical axis) at each value of $c \in \{0.5, 0.55, 0.6, \dots, 2\}$ (horizontal axis). Use abline() to add a vertical line at E[X].

Answer:

Question 2: Variance

Consider the random variable X characterized by the following PMF:

$$\begin{array}{c|cccc}
\hline
x & f(x) \\
\hline
1 & .2 \\
2 & .7 \\
3 & .1
\end{array}$$

(2a) Confirm that the variance of X is the same whether we compute it by the formula in Definition 2.1.12 or the Alternative Formula in Theorem 2.1.13. (We want the two variance computations, not the proof.)

Answer:

(2b) Write an R function to compute the variance of any discrete random variable. The arguments to your function should include the values the random variable can take on (x) and the probability it takes on each value (probs). Use your function to confirm your answer from question (2a).

Answer:

your code here

(2c) Explain what definition/property/mathematical operation is being used in each step of the following proof.

For random variable X and $a \in \mathbb{R}$, $V[aX] = a^2V[X]$,

$$V[aX] = E[(aX - E[aX])^2]$$
 (Step 1)

$$= E \left[(aX - aE[X])^2 \right]$$
 (Step 2)

$$= \operatorname{E}\left[a^{2}(X - \operatorname{E}[X])^{2}\right]$$
 (Step 3)

$$= a^2 \mathbf{E} \left[(X - \mathbf{E} [X])^2 \right] \tag{Step 4}$$

$$= a^2 V[X] \tag{Step 5}$$

Answer:

In Step $1, \ldots$

In Step $2, \ldots$

In Step $3, \ldots$

In Step $4, \ldots$

In Step $5, \ldots$

(2d) Use your variance function to confirm that $V[aX] = a^2V[X] \forall a \in \mathbb{R}$. (You can just show this for one value of a.)

Answer

(2e) Use your function to confirm that $V[X+c]=V[X] \,\forall\, c\in\mathbb{R}$. (You can just show this for one value of c.)

Answer