Problem set 3: Summarizing distributions

Due October 16, 2023, at 9pm

(Your name here)

NOTE: Start with the file `ps3_2023_summarizing_distributions.qmd` (available from the github repository at https://github.com/UChicago-pol-methods/IntroQSS-F23/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.

x	$f(x)$
-1	.3
2	.2
3	.5

(1a) Compute $\mathrm{E}[X]$. Show your work.

Answer:

$$\begin{aligned}\mathrm{E}[X] &= \sum_x x f(x) \\ &= -1 \times .3 + 2 \times .2 + 3 \times .5 \\ &= 1.6\end{aligned}$$

(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:

```
exp_func <- function(x, probs){
  sum(x * probs)
}
exp_func(x = c(-1, 2, 3), probs = c(.3, .2, .5))
```

```
[1] 1.6
```

(1c) Compute the MSE ($\mathrm{E}\left[(X-c)^2\right]$) for $c = 0$ and $c = 1$. Show your work.

Answer:

$$\begin{aligned}
\mathrm{E}\left[(X-0)^2\right] &= \sum_x (x-0)^2 f(x) \\
&= (-1)^2 \times .3 + (2)^2 \times .2 + (3)^2 \times .5 \\
&= 1 \times .3 + 4 \times .2 + 9 \times .5 \\
&= 5.6
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[(X-1)^2\right] &= \sum_x (x-1)^2 f(x) \\
&= (-1-1)^2 \times .3 + (2-1)^2 \times .2 + (3-1)^2 \times .5 \\
&= 4 \times .3 + 1 \times .2 + 4 \times .5 \\
&= 3.4
\end{aligned}$$

(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:

```
mse_func <- function(x, probs, c){
  sum(probs*(x - c)^2)
}
mse_func(x = c(-1, 2, 3), probs = c(.3, .2, .5), c = 0)
```

```
[1] 5.6
```

```
mse_func(x = c(-1, 2, 3), probs = c(.3, .2, .5), c = 1)
```

```
[1] 3.4
```

(1e) Create a vector `cs` that contains numbers in the sequence `1.0, 1.1, 1.2, . . ., 3`. 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`.

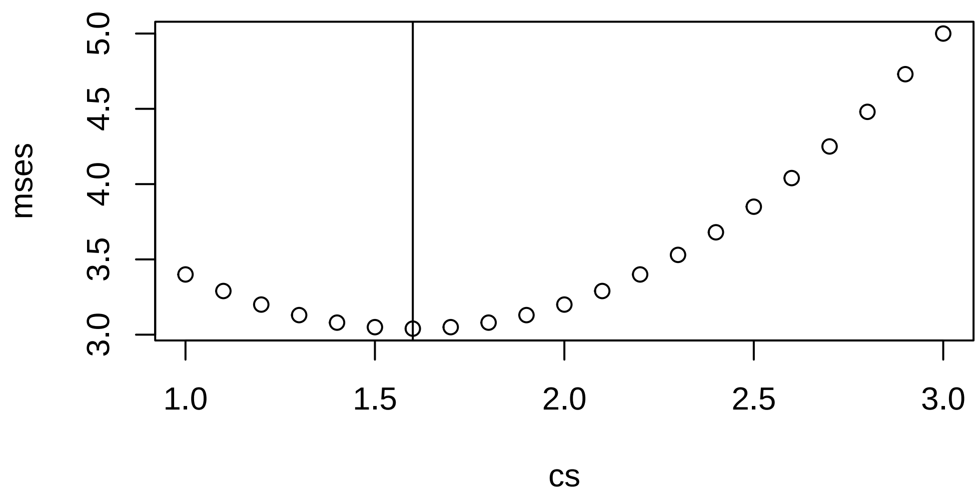
Answer:

```r
cs <- seq(from = 1, to = 3, by = .1)
mses <- rep(NA, times = length(cs))
for(i in 1:length(mses)){
  mses[i] <- mse_func(x = c(-1, 2, 3), probs = c(.3, .2, .5), c = cs[i])
}
```

(1f) Using the `plot()` command, make a scatterplot showing the MSE of X (vertical axis) at each value of $c \in \{1, 1.1, 1.2, \ldots, 3\}$ (horizontal axis). Use `abline()` to add a vertical line at $E[X]$.

Answer:

```r
plot(cs, mses)
abline(v = 1.6)
```



Question 2: Variance

Consider the random variable X characterized by the following PMF:

x	$f(x)$
0	.6
1	.1
2	.3

(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: First we calculate $\mathrm{E}[X]$:

$$\begin{aligned} \mathrm{E}[X] &= \sum_x x f(x) \\ &= 0 \times .6 + 1 \times .1 + 2 \times .3 \\ &= .7 \end{aligned}$$

Using Definition 2.1.12:

$$\begin{aligned} \mathrm{V}[X] &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{E}\left[(X - .7)^2\right] \\ &= (-.7)^2 \times .6 + .3^2 \times .1 + 1.3^2 \times .3 \\ &= .81 \end{aligned}$$

To use Theorem 2.1.13, we calculate $\mathrm{E}[X^2]$:

$$\begin{aligned} \mathrm{E}[X^2] &= \sum_x x^2 f(x) \\ &= 0 \times .6 + 1 \times .1 + 4 \times .3 \\ &= 1.3 \end{aligned}$$

Using Theorem 2.1.13:

$$\begin{aligned} \mathrm{V}[X] &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \\ &= 1.3 - .7^2 \\ &= .81 \end{aligned}$$

(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:

```r
# two versions, both of which use the expectation function I wrote above

# this version uses the E[(X - E[X])^2] formulation
var_func2 <- function(x, probs){
  ex <- exp_func(x, probs)
  exp_func((x - ex)^2, probs)
}

# this version uses the E[X^2] - E[X]^2 formulation
var_func <- function(x, probs){
  exp_func(x^2, probs) - exp_func(x, probs)^2
}

# confirming equality
var_func(x = c(0, 1, 2), probs = c(.6, .1, .3))
```

```
[1] 0.81
```

```r
var_func2(x = c(0, 1, 2), probs = c(.6, .1, .3))
```

```
[1] 0.81
```

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

For random variable X and $c \in \mathbb{R}$, $V[X+c] = V[X]$,

$$\begin{aligned}
\mathrm{V}[X+c] &= \mathrm{E}\left[(X+c-\mathrm{E}[X+c])^2\right] && \text{(Step 1)} \\
&= \mathrm{E}\left[(X+c-\mathrm{E}[X]-c)^2\right] && \text{(Step 2)} \\
&= \mathrm{E}\left[(X-\mathrm{E}[X])^2\right] && \text{(Step 3)} \\
&= \mathrm{V}[X] && \text{(Step 4)}
\end{aligned}$$

Answer:

In Step 1, . . .

In Step 2, . . .

In Step 3, . . .

In Step 4, . . .

Answer:

In Step 1, we use the definition of variance applied to the random variable $X + c$.

In Step 2, we use the linearity of expectations and the fact that the expectation of a constant c is c.

In Step 3, we simplify by observing that $c - c = 0$.

In Step 4, we use the definition of variance.

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

Answer

```
var_func(x = c(0, 1, 2), probs = c(.6, .1, .3))
```

```
[1] 0.81
```

```
var_func(x = c(0, 1, 2) + 3, probs = c(.6, .1, .3))
```

```
[1] 0.81
```

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

Answer

```
var_func(x = c(0, 1, 2), probs = c(.6, .1, .3))*9
```

```
[1] 7.29
```

```
var_func(x = c(0, 1, 2)*3, probs = c(.6, .1, .3))
```

```
[1] 7.29
```