STAT 639V - Assignment 1

Carson Stacy

9/1/2021

1. Let Y be a discrete random variable taking values from {1, 2, 3} with probabilities {0.2, 0.4, 0.4}. A sample of Y is {3, 2, 3, 1, 2, 2, 2, 3, 1}. Calculate the sample mean, sample variance, theoretical mean and theoretical variance for Y. (10 points)

given: $n = 9, y = \{3, 2, 3, 1, 2, 2, 2, 3, 1\}, P(1) = 0.2, P(2) = 0.4, P(3) = 0.4$

$\underline{Samplemean:}$

$$\bar{y} = \frac{1}{n} * \sum_{i=1}^{n} y_i = \frac{3+2+3+1+2+2+2+3+1}{9} = \frac{19}{9} = \boxed{2.\bar{1}}$$

$\underline{Samplevariance:}$

$$S^2 = \frac{1}{n-1} * \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{3(3-2.\bar{1})^2 + 4(2-2.\bar{1})^2 + 2(1-2.\bar{1})^2}{9-1} \approx \frac{2.37+0.05+2.47}{8} \approx \boxed{0.611}$$

$\underline{Theoreticalmean:}$

$$\mu = E(Y) = \sum_{j=1}^{k} y_j p_j = (1 * 0.2) + (2 * 0.4) + (3 * 0.4) = \boxed{2.2}$$

$\underline{Theoreticalvariance:}$

$$\sigma^2 = E(Y-\mu)^2 = \sum_{j=1}^{k} (y_j - \mu)^2 p_j = (1-2.2)^2 * 0.2 + (2-2.2)^2 * 0.4 + (3-2.2)^2 * 0.4 = \boxed{0.56}$$

2. The 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that (a) at least 2 are obese? (b) no more than 7 are obese? (10 points)

given:

$$n = 10, p = 0.262, P(k successes in n trials) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

(a) The probability that at least 2 are obese $P(k \geq 2)$ is equivalent to $1 - P(k < 2)$ which is equivalent to $P(k \geq 2) = 1 - [P(k = 0) + P(k = 1)]$

We can find that

$$P(k = 0) = \binom{n}{k} p^k (1-p)^{(n-k)} = \binom{10}{0} 0.262^0 (1 - 0.262)^{(10-0)} \approx 0.0479$$

and

$$P(k = 1) = \binom{n}{k} p^k (1-p)^{(n-k)} = \binom{10}{1} 0.262^1 (1 - 0.262)^{(10-1)} \approx 0.1701$$

which can then be used to calculate $P(k \geq 2)$:

$$P(k \geq 2) = 1 - [P(k = 0) + P(k = 1)] \approx 1 - (0.0479 + 0.1701) \approx \boxed{0.7819}$$

which can be confirmed in R code:

```r
pbinom(1,10,0.262, lower.tail = FALSE)
```

```
## [1] 0.781934
```

(b) The probability that no more than 7 are obese $P(k \leq 7)$ is equivalent to $1 - P(k > 7)$ which is equivalent to $P(k \leq 7) = 1 - [P(k = 8) + P(k = 9) + P(k = 10)]$

We can find that

$$P(k = 8) = \binom{n}{k} p^k (1-p)^{(n-k)} = \binom{10}{8} 0.262^8 (1 - 0.262)^{(10-8)} \approx 0.0005441712$$

and

$$P(k = 9) = \binom{n}{k} p^k (1-p)^{(n-k)} = \binom{10}{9} 0.262^9 (1 - 0.262)^{(10-9)} \approx 0.0000429307$$

and

$$P(k = 10) = \binom{n}{k} p^k (1-p)^{(n-k)} = \binom{10}{10} 0.262^{10} (1 - 0.262)^{(10-10)} \approx 0.000001524$$

which can then be used to calculate $P(k \leq 7)$:

$$P(k \leq 7) = 1 - [P(k = 8) + P(k = 9) + P(k = 10)] \approx 1 - (0.0005441712 + 0.0000429307 + 0.000001524) \approx \boxed{0.9994114}$$

which can be confirmed in R code:

```r
pbinom(7,10,0.262, lower.tail = TRUE)
```

```
## [1] 0.9994114
```

3. A bowl contains 10 marbles including 2 red marbles, 3 green marbles, and 5 blue marbles. If we randomly select 4 marbles from the bowl, WITH replacement, what is the probability of selecting 2 green marbles and 2 blue marbles? (10 points)

given: $n = 4, n_{red} = 2, n_{green} = 3, n_{blue} = 5, P(green) = 0.3, P(red) = 0.2, P(blue) = 0.5$

According to the multinomial formula,

$$P(n_{green} = 2, n_{blue} = 2, n_{red} = 0) = \frac{n!}{n_{green}!n_{blue}!n_{red}!} * P(green)^{n_{green}} * P(blue)^{n_{blue}} * P(red)^{n_{red}}$$

$$= \frac{4!}{2!2!0!} * 0.3^2 * 0.5^2 * 0.2^0 = 6 * 0.09 * 0.25 * 1 = \boxed{0.135}$$

4. Suppose X $\sim$ Pois(λ1) and Y $\sim$ Pois(λ2) are independent Poisson random variables. If λ1 = 2 and λ2 = 3, calculate E(X - Y) and Pr(X + Y = 1). (10 points)

Given $X \sim \mathsf{Pois}(\lambda_1)$ and $Y \sim \mathsf{Pois}(\lambda_2)$ where $\lambda_1 = 2$ and $\lambda_2 = 3$, the expectation of the difference of X and Y, $E(X - Y)$ is equal to $E(X) - E(Y)$ because the random variables are independent. For random variables following a poisson distribution, the $E(A) = \lambda_A$. In this case,

$$E(X - Y) = E(X) - E(Y) = \lambda_1 - \lambda_2 = 2 - 3 = \boxed{-1}$$

While λ of a poisson process is greater than zero $\lambda \in (0, \infty)$, the difference between two poisson processes, termed a Skellam Distribution, has a domain that includes negative integers, therefore I believe this answer is correct.

The $Pr(X + Y = 1)$ can be found according to the Probability Mass Function (PMF) for poisson processes below, where $j = 1$ and lambda of the linear combination of independent random variables X and Y of X + Y results in a $\lambda = 5$ because $E(X + Y) = E(X) + E(Y) = \lambda_1 + \lambda_2 = 2 + 3 = 5$

Poisson PMF

$$P(X + Y = j) = \frac{\lambda^j}{j!} * e^{-\lambda}$$

Given that $j = 1$ and $\lambda = 5$ for the linear combination X + Y, then

$$P(X + Y = 1) = \frac{5^1}{1!} * e^{-5} = 5e^{-5} \approx \boxed{0.03368}$$

5. Suppose random variable Y follows a normal distribution N(μ = 0, σ^2 = 4). Calculate the density at Y=2, i.e., f(2). Use R or table to calculate the inter-quartile range, Q.75 - Q.25.(10 points)

(a) Given $X \sim N(\mu = 0, \sigma^2 = 4)$, we know that $\sigma = \sqrt{4} = 2$.

From the PMF for the normal distribution

$$P(Y = y) = f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

we can find $P(y = 2)$:

$$P(Y = 2) = f(2) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(2-0)^2}{2*4}} \approx \boxed{0.1209854}$$

which can be confirmed in R:

```
dnorm(2, mean = 0, sd = 2)
```

```
## [1] 0.1209854
```

(b) The IQR of $X \sim N(\mu = 0, \sigma^2 = 4)$ can be determined by subtracting $Q_{0.75} - Q_{0.25}$, done here with using the qnorm function in R:

```
qnorm(p = 0.75, mean = 0, sd = 2) - qnorm(p = 0.25, mean = 0, sd = 2)
```

```
## [1] 2.697959
```

We can also see this via simulation

```
# set seed for reproducability
set.seed(1)
# create simulated dataset with mu=0 and sd=2, save as y
y <- rnorm(1000000, mean = 0, sd = 2)
#find the iqr for y
IQR(y)
```

```
## [1] 2.698783
```

```
# should be close to the theoretical IQR determined above..
```

$$IQR \approx \boxed{2.697959}$$

6. A coin was flipped 10 times independently, and 3 heads were observed. Let p be the probability of getting a head, compute the maximum likelihood estimate of p. (10 points)

given: $n = 10, y = 3, p = P(H)$

$$Y \sim \mathsf{Bin}(n, p)$$

The maximum likelihood estimate for p ($\hat{p}_{MLE}$) can be calculated according to $\hat{p}_{MLE} = \frac{y}{n}$ where n is the number of trials and y is the number of successes.

By the binomial proportion equation,

$$\hat{p}_{MLE} = \boxed{\frac{3}{10}}$$