Assignment 2: Binomial Distribution
Due September 19, 2023

John McCuan

corrected version (Problems 4 and 7) September 18, 2023

Problem 1 (simulation; from section 2.1.1 of my notes)

(a) Use a spreadsheet program (e.g., Libre Office Calc) to generate one hundred numbers $X_1, X_2, X_3, \ldots, X_{100}$ from among the numbers in the set $\{0, 1\}$ which "appear" to be chosen randomly according to the probability measure with $\beta(\{0\}) = 1/2$.

(b) Use the spreadsheet program to count how many "heads," i.e., $X_j = 1$, you obtained.

Problem 2 (binomial distribution; Problem 1 above)

(a) What is the probability that you did not get 50 for your answer to part **(b)** of Problem 1 above? Hint: What is the probability you did get 50?

(b) How many times would you expect to have to repeat the meaningful tasks associated with parts **(a)** and **(b)** of Problem 1 in order to see the answer "50" for (the execution of) part **(b)** at least once?

Problem 3 (Problem 2 above) If you calculated a probability in order to answer part **(b)** of Problem 2 above, what was the base model set S for your calculation? What is interesting about this set?

Problem 4 (simulation; from section 2.1.1 of my notes)

(a) Use a spreadsheet program (e.g., Libre Office Calc) to generate one hundred numbers $X_1, X_2, X_3, \ldots, X_{100}$ from among the numbers in the set $\{0, 1\}$ which "appear" to be chosen randomly according to the probability measure with $\beta(\{0\}) = 1/3$.

(b) Define a base set $S \subset \mathbb{R}^{100}$ to model the possible outcomes of the event that took place when you executed part **(a)** above.

(c) Define a probability measure $\pi : \wp(S) \to [0, 1]$ on S with

$$\pi\left(\left\{\sum_{j\in A} \mathbf{e}_j\right\}\right) = \left(\frac{2}{3}\right)^{\#A} \left(\frac{1}{3}\right)^{100-\#A}$$

for every $A \subset \{1, 2, 3, \ldots, 100\}$, and define a function $y : S \to \mathbb{N}$ by

$$y\left(\sum_{j\in A} \mathbf{e}_j\right) = \sum_{j\in A} j.$$

In these expressions $\mathbf{e}_j$ denotes the standard unit basis vector in $\mathbb{R}^{100}$. Find the value of the integral of the function y over S with respect to the measure π, and explain the meaning of this number.

Correction: The definition[1] of the function $y : S \to \mathbb{N}$ above is supposed to be

$$y\left(\sum_{j\in A} \mathbf{e}_j\right) = \sum_{j\in A} 1.$$

Addition: For those not familiar with the standard unit basis vectors, here is a little additonal explantion. The vector $\mathbf{e}_1 = (1, 0, 0, 0, \ldots)$ with 100 entries and the first one is 1. Similarly, $\mathbf{e}_2 = (0, 1, 0, 0. \ldots)$ with a 1 in the second component/entry. These can be added to get $\mathbf{e}_1 + \mathbf{e}_2 = (1, 1, 0, 0, 0, \ldots) \in \{0, 1\}^{100}$.

[1]The problem with the original definition of y makes sense, but I think it is very difficult to calculate the resulting integral by hand, and I don't see any natural explanation for the meaning of that integral. I think in a first version of the problem I intended for you to calculate the integral of the function originally given using mathematical software like Mathematica or a spreadsheet program. I decided to change the problem (making it both easier and more meaningful) but somehow I forgot to change the "j" to a "1." Sorry for the typo.

Problem 5 (sections 2.1.2 and 2.1.3 in my notes) Let $\beta : \mathcal{P}(\mathbb{R}) \to [0, 1]$ denote the Bernoulli measure. Let $x : \mathbb{R} \to \mathbb{R}$ by $x(t) = t + 5$ be a renaming bijection.

(a) Find the measure induced by the renaming x.

(b) Find the PMF of the induced measure obtained in part **(a)** above. How is this PMF different from the PMF of the Bernoulli measure?

Problem 6 (section 2.1.3 and 2.1.4 in my notes, especially Exercise 2.1.8; also see the discussion of Orloff and Booth concerning intransitive dice) Let

$$S = \{\text{one}, \text{two}, \text{three}, \text{four}, \text{five}, \text{six}\}.$$

Consider the uniform probability measure π on S. Now, say you have three (standard six sided) dice—well, not completely standard: One is red and has five sides marked "three," i.e., with three dots, and one side marked "six." A second die is green and has one side marked "one" and five sides marked "four." The third die is blue with three sides marked "two" and three sides marked "five."

(a) Define three real valued functions $r : S \to \mathbb{R}$, $g : S \to \mathbb{R}$ and $b : S \to \mathbb{R}$ appropriate for modeling the possible outcomes of rolling the red, green, and blue die respectively.

(b) Find the induced measures α_r, α_g and α_b associated with each of the functions from part **(a)** above and plot the associated PMF.

(c) Find the integrals of r, g and s with respect to π.

Problem 7 (Problem 6 above) Consider the red, green and blue dice of Problem 6 above. Let

$$C = S \times S = \{(\omega_1, \omega_2) : \omega_j \in S\}$$

be a set used to model the outcome of a competition between a pair of the three dice from Problem 6.

(a) Describe the uniform measure on C.

(b) What function $h : C \to \mathbb{R}^2$ is appropriate to model the outcome of a competition between the red die and the blue die?

(c) Using your function from part **(b)** find the measure α_w induced on C by $w = \delta \circ h : C \to \mathbb{R}$ where $\delta(\xi_1, \xi_2) = \xi_2 - \xi_1$.

Correction: An induced measure measures the subsets in the codomain of the inducing function. In this case, the induced measure should be a measure on $\mathbb{R}$, so if I want this to make sense, I should write

> Using your function from part **(b)** find the measure α_w induced on $\mathbb{R}$ by $w = \delta \circ h : C \to \mathbb{R}$ where $\delta(\xi_1, \xi_2) = \xi_2 - \xi_1$.

Having made this correction, let me mention something: The induced measure involves preimages under w back in the set C, and I think it is important to see those preimages back in C and the corresponding values of the base (uniform) measure. That's what I had in mind. Perhaps I should add an additional part making this explicit, but I'll leave it for now.

(d) Using your induced measure $\alpha_w : \wp(C) \to [0, 1]$ what is the value and meaning of

$$\alpha_w \{ \{(\omega_1, \omega_2) : w(\omega_1, \omega_2) > 0\} \}? \tag{1}$$

Correction: This part[2] should read

> Using your induced measure $\alpha_w : \wp(\mathbb{R}) \to [0, 1]$ what is the value and meaning of
>
> $$\alpha_w(\{\xi \in \mathbb{R} : \xi > 0\})?$$

[2]if I want it to be correct and make sense

You'll note that there is also an incorrect grouping symbol in (1); the first "curly bracket" should just be round evaluation parentheses. Incidentally, the set

$$\{(\omega_1, \omega_2) \in C : w(\omega_1, \omega_2) > 0\}$$

appearing in (1) is essentially the set I'd like to make sure you "see."

(e) Repeat parts **(b)** and **(c)** for competitions between the remaining pairs of dice. What interesting thing do you find?

Problem 8 (From sections 2.4.1 and 2.4.2 in my notes)

(a) Plot the PMF of the binomial distribution when $p = 1$ and $n = 1, 2, 3, \ldots$.

(b) Plot the CMF of the binomial distribution when $p = 1/2$ and $n = 2, 3, 4$.

Problem 9 (binomial distribution) Let β denote the binomial measure and α_x the binomial induced measure.

(a) Calculate

$$\lim_{p \searrow 0} \alpha_x(\{0\}) \qquad \text{and} \qquad \lim_{p \searrow 0} \alpha_x(\{1\}).$$

(b) Calculate

$$\lim_{p \nearrow 1} \alpha_x(\{0\}) \qquad \text{and} \qquad \lim_{p \nearrow 1} \alpha_x(\{1\}).$$

(c) Find the unique value p_0 of p for which $\alpha_x(\{0\}) = \alpha_x(\{1\})$.

(d) What is the relation between $\alpha_x(\{0\})$ and $\alpha_x(\{1\})$ when $p < p_0$?

(d) What is the relation between $\alpha_x(\{0\})$ and $\alpha_x(\{1\})$ when $p > p_0$?

Problem 10 (binomial distribution; Problem 9 above) Let β denote the binomial measure and α_x the binomial induced measure.

(a) Find the unique value p_* of p for which

$$\alpha_x(\{0\}) = \sum_{j=1}^{n} \alpha_x(\{j\}). \tag{2}$$

(b) If p is the probability of "success" in an event modeled by the Bernoulli base set $\{h, t\}$ and $\alpha_x(\{\xi\})$ is the probability of having ξ "successes" out of n Bernoulli trials, describe the probabilistic interpretation of the inequality $p < p_*$.