

# **GV300 - Quantitative Political Analysis**

University of Essex - Department of Government

---

Lorenzo Crippa

Week 7 – 11 November, 2019

Week 8: Office hour on Monday, from 14 to 16  
(**NOT** Wednesday)

Plus: From now on, extra support hour dedicated to problem sets  
on Monday from 14 to 16 every week a problem set is due

Office 5B.153  
l.crippa@essex.ac.uk

## Problem Set 2.1

Events  $E$ ,  $F$ , and  $G$  form a list of three mutually exclusive, collectively exhaustive events with  $P(E) \neq 0$ ,  $P(F) \neq 0$ , and  $P(G) \neq 0$ . Determine for each of the following statements whether it must be true, could be true, or must be false. Show your reasoning (9 marks, 3 each).

## Problem Set 2.1 – Solution

- (a)  $E'$ ,  $F'$ , and  $G'$  are mutually exclusive. False.  $E' = F \cup G$   
 $F' = E \cup G$  and  $G' = E \cup F$
- (b)  $E'$ ,  $F'$ , and  $G'$  are collectively exhaustive. True. They cover the full sample space
- (c)  $E'$  and  $F'$  are independent. False, the fact that  $E' = F \cup G$  occurs affects the probability that  $F' = E \cup G$  occurs.

## Problem Set 2.2

Say, in the UK, 38% of citizens support the Conservatives, 33% support Labour, and 29% support the LibDems. In the last election, 55% of supporters of the Conservatives voted, 43% of Labour supporters voted, and 59% of LibDem supporters voted. What is the probability that a citizen voted? (6 marks)

## Problem Set 2.2 – Solution

$$\begin{aligned}P(V) &= P(V \cap C) + P(V \cap L) + P(V \cap LD) = \\&P(V|C)P(C) + P(V|L)P(L) + P(V|LD)P(LD) = \\&.55 * .38 + .43 * .33 + .59 * .29 = .522\end{aligned}$$

## Problem Set 2.3

if  $P(A) = 0.3$ ,  $P(B') = 0.65$ , and  $P(A \cup B) = 0.65$ , determine the following (and show the steps of your computation, 12 marks, 4 each)

- (a)  $P(B)$
- (b)  $P(A \cap B)$
- (c)  $P(A'|B')$

## Problem Set 2.3 – Solution

if  $P(A) = 0.3$ ,  $P(B') = 0.65$ , and  $P(A \cup B) = 0.65$ , determine the following (and show the steps of your computation, 12 marks, 4 each)

(a)  $P(B) = 0.35$  from  $P(B') = 0.65$

(b) we know  $P(A \cup B) = 0.65$ ,  $P(A) = 0.3$ ,  $P(B) = 0.35$ . We also know  $P(A \cup B) = P(A) + P(B)$  only if no overlap of A and B. Thus  $P(A \cap B) = 0$

(c)  $P(A'|B') = \frac{P(A' \cap B')}{P(B')} = \frac{0.35}{0.65} = .54$  because  $P(A' \cap B') = 1 - P(A \cup B)$ , that is, every outcome not in the union of A and B is in the intersection of  $A'$  and  $B'$ .



## Problem Set 2.4

Let  $D$  be the event that a fetus has Down syndrome and  $B$  be the event that a blood test screening for this kind of birth defect is positive. Further, assume that  $p(B|D) = 0.96$ ,  $p(B'|D') = 0.95$ , and  $p(D) = 0.002$  (these numbers are roughly accurate; 10 marks).

- (a) Interpret  $p(B|D)$ ,  $p(B'|D')$  and  $p(D)$  in words (4 marks).
- (b) Solve for  $p(D|B)$  and interpret this probability in words (6 marks).

## Problem Set 2.4 – Solution

(a) Interpret  $p(B|D)$ ,  $p(B'|D')$  and  $p(D)$  in words (4 marks).

- $p(B|D)$  is the probability of the blood test being positive given that the fetus has Down Syndrome. That is in 96% of the tests.
- $p(B'|D')$ : There is a .95 probability that the blood test is negative if the fetus does not have Down Syndrome
- $p(D)$ : There is a .002 probability that the fetus has Down Syndrome.

# Bayes' Theorem

Let  $A_1, \dots, A_n$  be a partition of the sample space and  $B$  an event associated with the sample space.

Bayes' Theorem:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}, i = 1, \dots, n$$

## Problem Set 2.4 – Solution

(b) Solve for  $p(D|B)$  and interpret this probability in words (6 marks).

- We know that  $p(B|D) = 0.96$ ,  $p(B'|D') = 0.95$  and  $p(D) = 0.002$
- $p(B'|D') = 1 - p(B|D')$ , then  $p(B|D') = .05$
- $D$  and  $D'$  are partitions of the sample space.

$$P(D|B) = \frac{p(B|D)p(D)}{p(B|D)p(D) + p(B|D')p(D')}$$

$$P(D|B) = \frac{0.96 * 0.002}{0.96 * 0.002 + 0.05 * 0.998} = 0.037$$

- The probability that the fetus has Down syndrome given that the blood test comes back positive is 0.037

## Problem Set 2.5

Define the random variables  $H$  “Number of heads in 3 tosses” with realization  $h^0$  and  $R$  “The length of runs of the same outcome, heads or tails, in 3 tosses” with realization  $r^0$ . For example, the outcome “HHT” would be scored as  $H = 2$  and  $R = 2$  because it contains 2 heads and a length of a run of heads of 2 (2 heads right after each other in the first two tosses) (14 marks).

- (a) Reproduce the probability mass function of  $H$
- (b) Show the probability mass function of  $R$
- (c) Compute and table the joint probabilities of those realizations of the random variables  $H$  and  $R$  that may occur jointly. Also add the marginal probabilities to the table

## Problem Set 2.5 – Solution

- (a) For each realization of the random variable  $H = h^0$ , its probability will be:

$$P(H = h^0) = \binom{n}{h^0} p^{h^0} (1-p)^{n-h^0} = \frac{n!}{h^0!(n-h^0)!} p^{h^0} (1-p)^{n-h^0}$$

Where  $n = 3$  and  $p = .5$

$H$  can take values  $H = \{0, 1, 2, 3\}$ . Substitute each of these values in the above formula and you get:

$$P(H = 0) = 1/8, P(H = 1) = 3/8, P(H = 2) = 3/8,$$

$$P(H = 3) = 1/8$$

- (b) Tossing a coin three times can result in the 8 following events:  $TTT, TTH, THT, THH, HHH, HHT, HTH, HTT$ . Thus  $R$  can take values in the set  $\{1, 2, 3\}$  with the following probabilities:

$$P(R = 1) = 2/8 = 1/4$$

$$P(R = 2) = 4/8 = 1/2$$

$$P(R = 3) = 2/8 = 1/4$$

## Problem Set 2.5 – Solution

- (c) 1. How many ways are there to obtain  $H = 0$  and  $R = 1$  at the same time? What is the joint probability of these two events given that there are 8 possible outcomes? You can build the table below on joint probabilities asking these questions:

Definition of joint probability:

$$P_{H,R}(h_0, r_0) = P_{H,R}(r_0|h_0)P_H(h_0) = P_{H,R}(h_0|r_0)P_R(r_0)$$

		$H$				
		0	1	2	3	$p(r_0)$
$R$	1	0	1/8	1/8	0	1/4
	2	0	1/4	1/4	0	1/2
	3	1/8	0	0	1/8	1/4
$p(h_0)$		1/8	3/8	3/8	1/8	1



## Problem Set 2.6

---

Compute the PMF, mean, and variance of the following random variable and show your computations:  $y$  is the number of successful applications of economic sanctions in 7 countries where the probability that economic sanctions are successful is .3 (10 marks)

## Problem Set 2.6 – Solution

Use the binomial PMF:

$$P(y = 0) = (1).7^7 = .082$$

$$P(y = 1) = (7).3^1.7^6 = .247$$

$$P(y = 2) = (21).3^2.7^5 = .312$$

$$P(y = 3) = (35).3^3.7^4 = .227$$

$$P(y = 4) = (35).3^4.7^3 = .097$$

$$P(y = 5) = (21).3^5.7^2 = .025$$

$$P(y = 6) = (7).3^6.7^1 = .004$$

$$P(y = 7) = (1).3^7 = .0002;$$

$$E(y) = n * p = 7 * 0.3 = 2.1$$

$$V(y) = n * p * (1 - p) = 7 * .3 * .7 = 1.47$$

Or you use the general formulas for the weighted mean and variance

## Problem Set 2.7

---

Using your preferred statistical software, demonstrate that your calculations in problem 7 are approximately correct. Please turn in your run file (or copy and paste the code you wrote into your answer sheet) as well as any plot or table you produced to answer this question (10 marks)

## Problem Set 2.7 – Solution

Solution in R:

```
1 y <- rbinom(1000, size = 7, p = 0.3)
2
3 hist(y)
4
5 mean(y) # 2.205, very close to 2.1
6 var(y) # 1.49, very close to 1.47
```

Solution in Stata:

```
1 set obs 1000
2 gen y = rbinomial(7, 0.3)
3
4 histogram y
5 sum y
6 * std deviation is 1.208187 thus the variance is:
7 display 1.208187^2
```

## Problem Set 2.8

---

Consider this case: chances are, Boris Johnson will be battling parliament to get Brexit legislation passed for some time. Say, he has to bring four bills to the floor to get the necessary Brexit legislation passed but even Tories are skeptical of his leadership. They say, “he is a terrible leader who loses in parliament too often.” Suppose Tory backbenchers’ interpretation of “losing too often” and therefore being a terrible leader is to lose 7 out of 10 votes submitted to parliament. Further suppose that Johnson gets 3 out of 4 bills voted into law eventually. Thinking about this hypothetical case, please answer all of the following questions (12 marks, 4 marks each):

## Problem Set 2.8

---

- (a) what in the description above is the data we would observe?
- (b) what would be the random variable we could build to study this case?
- (c) what would be a theoretical claim we could test using this collected data? Formulate a testable null hypothesis. What would be your alternative hypothesis?

## Problem Set 2.8 – Solution

$X$ : “number of votes he loses” indicates whether he is a terrible leader. Binomial choice: The success probability of the event “lose a vote” is  $p = .7$ . What is the probability now that we observe  $X = 0, 1, 2, 3, 4$  lost votes out of 4 attempts? Apply the binomial PMF:

$$P(X = 0) = .0081, P(X = 1) = .0756, P(X = 2) = .2646, \\ P(X = 3) = .4116, P(X = 4) = .2401$$

Hypothesis: “Johnson is a terrible leader”.  $H_0$  “The probability of losing a vote is .7”  $H_A$  “The probability of losing a vote is lower than .7” We observe voting record and we test whether we can sustain that the probability with which he loses votes is, indeed, .7.

## Problem Set 2.9

Donald Trump's cabinet is an unruly bunch. Let's say we record the number of departures (resignations) from his cabinet on a monthly basis. The table below gives the number of departures per month and the associated probability distribution (10 marks):

Number of departures	0	1	2	3	4	5
Probability	.5	.24	.15	.07	.03	.01

Calculate mean and variance of the number of departures.



## Problem Set 2.9 – Solution

Number of departures	0	1	2	3	4	5	$\Sigma$
Probability	.5	.24	.15	.07	.03	.01	1
$xp(x)$	0	.24	.3	.21	.12	.05	.92
$x^2p(x)$	0	.24	.6	.63	.48	.25	2.2

$$E(X) = .92 \text{ and } V(X) = 2.2 - .92^2 = 1.35$$

## Math Refresher - Exercises in R and Stata

Compute the following (using either R/Stata or just pen and paper):

1.

$$\begin{bmatrix} 2 & 0 & -4 \\ 1 & 3 & 5 \\ -3 & 1 & 4 \\ 1 & 2 & 2 \end{bmatrix} \times \begin{bmatrix} 1 & -4 \\ -1 & 2 \\ 3 & 5 \end{bmatrix}$$

2.

$$\begin{bmatrix} 2 & 0 & -4 \\ 1 & 3 & 5 \\ -3 & 1 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & -4 & 1 \\ -1 & 2 & 3 \end{bmatrix}$$

3.

$$\begin{bmatrix} 2 & 0 & -4 \\ 1 & 3 & 5 \\ -3 & 1 & 4 \end{bmatrix} \times \begin{bmatrix} 2 & 1 & -3 \\ 0 & 3 & 1 \\ -4 & 5 & 4 \end{bmatrix}$$

## Math Refresher - Exercises in R and Stata

Generate a vector of 1000 random numbers (it should contain at least a 0) and call it *x*. Obtain the functions below and plot them in twoway graphs. Then calculate their derivatives and plot them on a twoway graph with the original functions.

1. Linear

$$y = 3x - 4$$

2. Quadratic

$$y = -x^2 + 3x - 4$$

3. Cubic

$$y = x^3 - x^2 + 3x - 4$$

4. Logarithm

$$y = \ln(x)$$

5. Exponential

$$y = e^{(x)}$$

6. Trigonometric

$$y = \sin(x) + 1.3$$

# Loops, functions and programs in R and Stata

## Exercises for R users:

1. A function for the median and the standard deviation
  - a. re-program a function for median and population standard deviation and call them “median2” and “sd2”
  - b. generate two vectors (of even and uneven length), apply the new functions to them and compare them with those obtained by applying base R's functions.
2. Imagine you're tossing a coin  $n$  times. ( $X$ : number of heads)
  - a. write a function that returns a data frame with all  $X$ , all possible ways to get  $X$ , and the probability of each  $X$
  - b. apply the function to  $n = 10$ ,  $p = 0.5$ . Generate a twoway plot (with  $X$  on the x-axis and  $p$  on the y-axis).
  - c. Draw 1000 observations for the number of heads obtained by tossing 10 times a fair coin. Obtain a histogram reporting the results and compare it with the plot from b.

Exercises for Stata users:

1. Program a function that takes as arguments:
  - a. the number of observations to be drawn from a random binomial distribution
  - b. the number of trials
  - c. the probability of success for a trial
2. Simulate the following:
  - a. toss a fair coin 2, 20, 200, 2000, 4000 and 8000 times and compute the mean number of heads
  - b. store a boxplot and a histogram relative to the means obtained for each of the six iterations and export a pdf showing them side by side
  - c. export a graph showing all iterations and all plots together (a total of 12 plots, arranged in 3 rows)

# Median, mean and population standard deviation

**Median:** “it is the point such that as many cases are greater as are less” (Gill 2006, 362). What if the variable is even in length?

**Mean:**

$$\mu_X = \frac{1}{N} \sum_{i=1}^N X_i$$

**Population standard deviation:**

$$\sigma_X = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu_X)^2}$$

# The binomial distribution

**Number of possible ways each outcome can occur:**

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

**Probability that a certain outcome occurs:**

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

All clear? Questions?  
Thanks and see you next week!



Gill, Jeff (2006). *Essential Mathematics for Political and Social Research*. Cambridge University Press.