STAT 35A HW2 Solutions

http://www.stat.ucla.edu/~dinov/courses_students.dir/09/Spring/STAT35.dir

1. A computer consulting firm presently has bids out on three projects. Let A_i = { awarded project i }, for i = 1, 2, 3, and suppose that $P(A_1) = 0.22$, $P(A_2) = 0.25$, $P(A_3) = 0.28$, $P(A_1 \cap A_2) = 0.11$, $P(A_1 \cap A_3) = 0.05$, $P(A_2 \cap A_3) = 0.07$, $P(A_1 \cap A_2 \cap A_3) = 0.01$. Express in words each of the following events and compute the probability of each event.

1.A (4 POINTS): $A_1 \cup A_2$

The probability that either $\mathbf{A_1}$ or $\mathbf{A_2}$ occurs:
$P(\mathbf{A_1}) + P(\mathbf{A_2}) - P(\mathbf{A_1} \cap \mathbf{A_2}) = .22 + .25 - .11 = .36$

1.B (4 POINTS): $A'_1 \cap A'_2$

The probability that neither $\mathbf{A_1}$ or $\mathbf{A_2}$ occurs:

$P(\mathbf{A'_1} \cap \mathbf{A'_2}) = P(\mathbf{A_1} \cup \mathbf{A_2})' = 1-P(\mathbf{A_1} \cup \mathbf{A_2}) = 1-.36 = .64$

1.C (4 POINTS): $A_1 \cup A_2 \cup A_3$

The probability that either $\mathbf{A_1}$ or $\mathbf{A_2}$ or $\mathbf{A_3}$ occurs:
$P(\mathbf{A_1})+ P(\mathbf{A_2})+ P(\mathbf{A_3})- P(\mathbf{A_1} \cap \mathbf{A_2})- P(\mathbf{A_2}\cap \mathbf{A_3})- P(\mathbf{A_1} \cap \mathbf{A_3})+ P(\mathbf{A_1} \cap \mathbf{A_2} \cap \mathbf{A_3})$
$= 0.22+0.25+0.28-0.11-0.07-0.05+0.01 = .53$

1.D (4 POINTS): $A'_1 \cap A'_2 \cap A'_3$

The probability that neither $\mathbf{A_1}$ or $\mathbf{A_2}$ or $\mathbf{A_3}$ occurs:

$=P(\mathbf{A_1} \cup \mathbf{A_2} \cup \mathbf{A_3})' = 1 – P(\mathbf{A_1} \cup \mathbf{A_2} \cup \mathbf{A_3}) = 1-.53 = .47$

1.E (5 POINTS): $A'_1 \cap A'_2 \cap A_3$

The probability that neither $\mathbf{A_1}$ or $\mathbf{A_2}$ occurs, and $\mathbf{A_3}$ occurs:
Observing that $P(X \cap Y)=P(X)-P(X \cap Y')$, and using $X=A'_1 \cap A'_2$ and $Y=A_3$ we get

$P(\mathbf{A_1} \cup \mathbf{A_2} \cup \mathbf{A_3}) - P(\mathbf{A_1} \cup \mathbf{A_2}) = .53 - .36 = .17$

You can also reason about this graphically by drawing a Venn diagram to get the same answer.

1.F (4 POINTS): $(A'_1 \cap A'_2) \cup A_3$

The probability that neither $\mathbf{A_1}$ or $\mathbf{A_2}$ occurs, or $\mathbf{A_3}$ occurs:
Using the additive rule, and the results from 1.E, we get
$P(\mathbf{A'_1} \cap \mathbf{A'_2}) + P(\mathbf{A_3}) - P(\mathbf{A'_1} \cap \mathbf{A'_2} \cap \mathbf{A_3}) = .64 + .28 - .17 = .75$

2. An insurance company offers four different deductible levels – none, low, medium, and high – for its homeowner's policyholders and three different levels – low (L), medium (M), and high (H) – for its automobile policyholders. The accompanying table gives proportions for the various categories of policyholders who have both types of insurance. For example, the proportion of individuals with both low homeowner's deductible and low auto deductible is 0.06 (6% of all such individuals). Suppose an individual having both types of policies is randomly selected.

		Home policies			
		None	Low	Medium	High
Auto Policies	Low	0.04	0.06	0.05	0.03
	Medium	0.07	0.10	0.20	0.10
	High	0.02	0.03	0.15	0.15

2.A (4 POINTS): WHAT IS THE PROBABILITY THAT THE INDIVIDUAL HAS A MEDIUM AUTO DEDUCTIBLE AND A HIGH HOMEOWNER'S DEDUCTIBLE?

"And" corresponds to the intersection of the high home column and the medium auto row, which is 0.10.

2.B (4 POINTS): WHAT IS THE PROBABILITY THAT THE INDIVIDUAL HAS A LOW AUTO DEDUCTIBLE? A LOW HOMEOWNER'S DEDUCTIBLE?

Summing the entire row for low auto: 0.04+0.06+0.05+0.03 = .18
Summing the entire column for low home: 0.06 + 0.10 + 0.03 = .19

2.C (4 POINTS): WHAT IS THE PROBABILITY THAT THE INDIVIDUAL IS IN THE SAME CATEGORY FOR BOTH AUTO AND HOMEOWNER'S DEDUCTIBLES?

Sum the diagonal where both home and auto policies are low, medium, and high: 0.06 + 0.20 + 0.15 = .41

2.D (4 POINTS): BASED ON YOUR ANSWER IN PART (C), WHAT IS THE PROBABILITY THAT THE TWO CATEGORIES ARE DIFFERENT?

This is the compliment of the event defined in (c): 1-.41 = .59

2.E (5 POINTS): WHAT IS THE PROBABILITY THAT THE INDIVIDUAL HAS AT LEAST ONE LOW DEDUCTIBLE LEVEL?

Sum both the low auto row and low home column but don't double count the intersection: (0.04+0.06+0.05+0.03) + (0.06+.10+.03) – 0.06 = .31

2.F (4 POINTS): USING THE ANSWER IN PART (E), WHAT IS THE PROBABILITY THAT NEITHER DEDUCTIBLE LEVEL IS LOW?

Again, taking the compliment of (e): 1-.31 = .69

3. **A card is drawn from a standard well-shuffled deck of 52 playing cards. What is the theoretical probability of the event A = {the drawn card is a king or a club}? In this applet an outcome of a King is recorded in the first variable Y and corresponds to Y=13; the second variable Z corresponds to the suit of the card with a Club, ♣, represented by Z=0. Using this Card-Experiment, from SOCR Experiments, perform 20 experiments and determine the proportion of the outcomes the event A was observed. This would correspond to the empirical probability (the chance) of the event A. How close are the observed and the theoretical probabilities for the event A? Would the discrepancy between these increase or decrease if the number of hands drawn increases? Experiment and report!**

3.A WHAT IS THE PROBABILITY OF DRAWING A KING OR A CLUB?

P(king) = 1/13
P(club) = 1/4
P(king ∩ club) = 1/52

Therefore P(king ∪ club) = P(king)+P(club)- P(king ∩ club) = 1/13 + 1/4 -1/52 = .3077

3.B WHAT IS THE EMPIRICAL PROBABILITY OF EVENT A USING YOUR 20 TRIALS?

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Denom	6	4	11	13	10	6	12	1	11	7	10	8	1	11	9	10	3	5	12	8
Suit	2	1	0	1	1	1	1	2	1	3	2	3	3	2	2	3	0	3	0	1

In my 20 trials, I got 4 outcomes that were either a king or a club. This corresponds to an empirical probability of 4/20=0.2.

3.C HOW CLOSE ARE THE OBSERVED AND THEORETICAL PROBABILITIES FOR EVENT A? WOULD THE DISCREPANCY BETWEEN THESE INCREASE OR DECREASE IF THE NUMBER OF HANDS DRAWN INCREASES?

My discrepancy was .1077. We would expect this to decrease as we draw more hands.

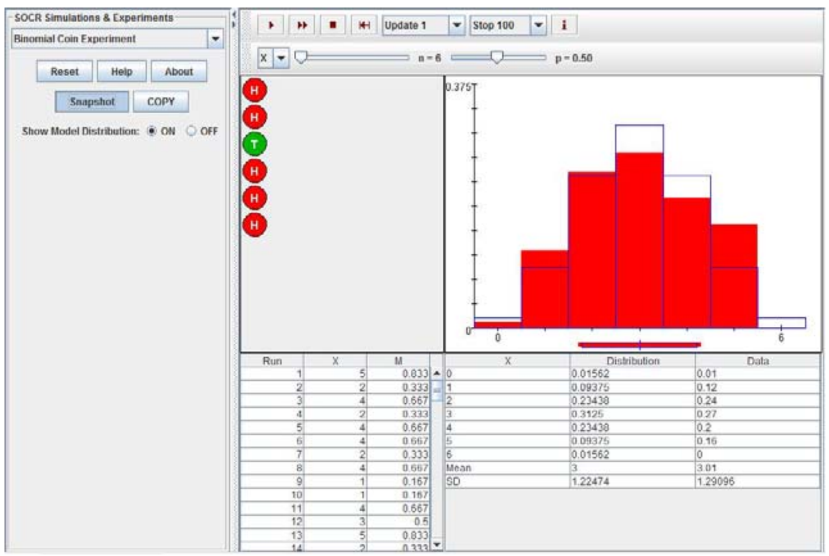
4. **Suppose we need to validate that a coin given to us is fair. We toss the coin 6 times independently and observe only one Head. If the coin was fair P(Head)=P(Tail)=0.5 and we would expect about 3 Heads and 3 Tails.**

4.A (10 POINTS): UNDER THESE FAIR COIN ASSUMPTIONS WHAT IS THE (THEORETICAL) PROBABILITY THAT ONLY 1 HEAD IS OBSERVED IN 6 TOSSES?

Using the binomial distribution with p=0.5 (fair coin), n=6 (total flips), and k=1 (number of heads) we have:

$$\binom{n}{k}p^k(1-p)^{n-k} = \frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k} = \frac{720}{120}(0.5)^1(1-0.5)^5 = 6 \cdot 0.5 \cdot .03125 = .09375$$

4.B (7 POINTS): EMPIRICALLY COMPUTE THE ODDS (CHANCES) OF OBSERVING ONE HEAD IN 6 FAIR-COIN-TOSSES (RUN 100 EXPERIMENTS AND RECORD THE NUMBER OF THEM THAT CONTAIN EXACTLY 1 HEAD)

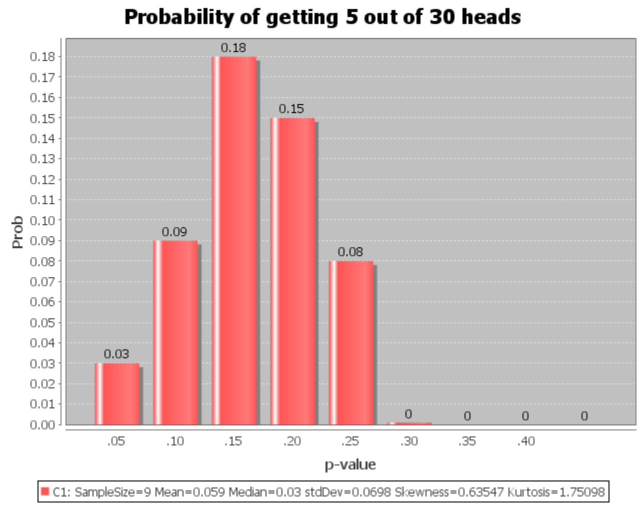


Running 100 coin flip experiments, 1 coin was heads in 12 of the trials for a frequency of .12.

4.C (8 POINTS): EMPIRICALLY ESTIMATE THE BIAS OF THE COIN WE HAVE TESTED. EXPERIMENT WITH TOSSING 30 COINS AT A TIME. YOU SHOULD CHANGE THE P-VALUE=P(HEAD), RUN EXPERIMENTS AND PICK A VALUE ON THE X-AXIS THAT THE EMPIRICAL DISTRIBUTION (RED-HISTOGRAM) PEAKS AT. PERHAPS YOU WANT THIS PEAK X VALUE TO BE CLOSE TO THE OBSERVED 1-OUT-OF-6 HEAD-COUNT FOR THE ORIGINAL TEST OF THE COIN. INCLUDE SOME GRAPH SNAPSHOTS AND SOME BRIEF DISCUSSION REGARDING YOUR EXPERIMENTS AND FINDINGS.

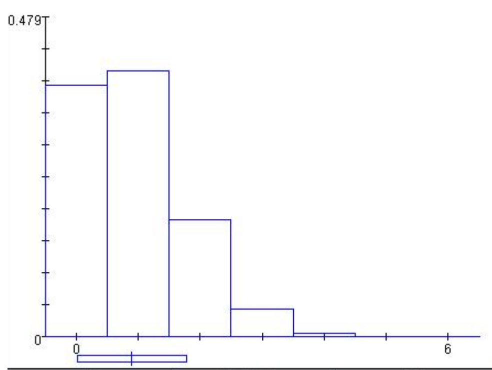
Below are my results of getting 5 heads out of 30 after running 100 trials for a range of p-values:

p-value	Relative frequency of getting 1 head
.05	.030
.10	.090
.15	.180
.20	.150
.25	.080
.30	.001
.35	.000
.40	.000



Empirically, the p-value that gives the highest frequency of getting 5 head out of 30 is .15. If I plot the results of my trials as a histogram, it is clear that this that the peak is indeed around .15.

You can use the binomial coin experiment to visualize the true distribution where you'll see that the probability of getting 1 head out of 6 is indeed maximized around a p-value of .15.



X	Distribution	Data
0	0.37715	
1	0.39933	
2	0.17618	
3	0.04145	
4	0.00549	
5	0.00039	
6	0.00001	
Mean	0.9	
SD	0.87464	

Plot of binomial distribution with p=.15 and n=6

5.

5.A (6 POINTS): IN THE BALL AND URN EXPERIMENT HOW DOES THE DISTRIBUTION OF THE NUMBER OF RED BALLS (Y) DEPEND ON THE SAMPLING STRATEGY (WITH OR WITHOUT REPLACEMENT)?

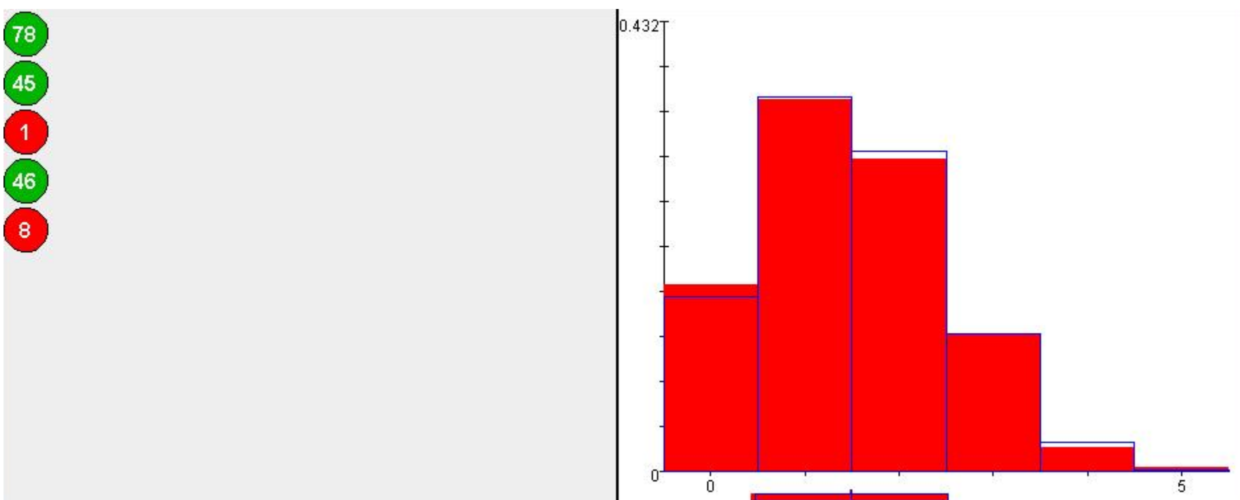
At one extreme, if you only sample one ball then replacing it has no impact on your sample. In this respect, the distributions are identical. At the other extreme, where n=N (sample size is the same as the total number of balls), if you are not replacing balls the only possible outcome is the exact ratio of red balls because you removed every ball from the urn. If n is somewhere in the middle between 0 and N, the variance of your outcomes will be smaller when not using replacement. The replacement strategy never affects the mean.

5.B (6 POINTS): DO N, R AND N ALSO PLAY ROLES?

Yes, for any given N and R, the influence of not using replacement (decrease in variance) is stronger as n increases.

5.C (6 POINTS): SUPPOSE N=100, n=5, R=30 AND YOU RUN 1,000 EXPERIMENTS. WHAT PROPORTION OF THE 1,000 SAMPLES HAD ZERO OR ONE RED BALLS IN THEM? RECORD THIS VALUE.

Using replacement, these are the results of my experiment. The number of balls with 0 or 1 are 1000*(.18+.358)=538 according to the data column of the output.

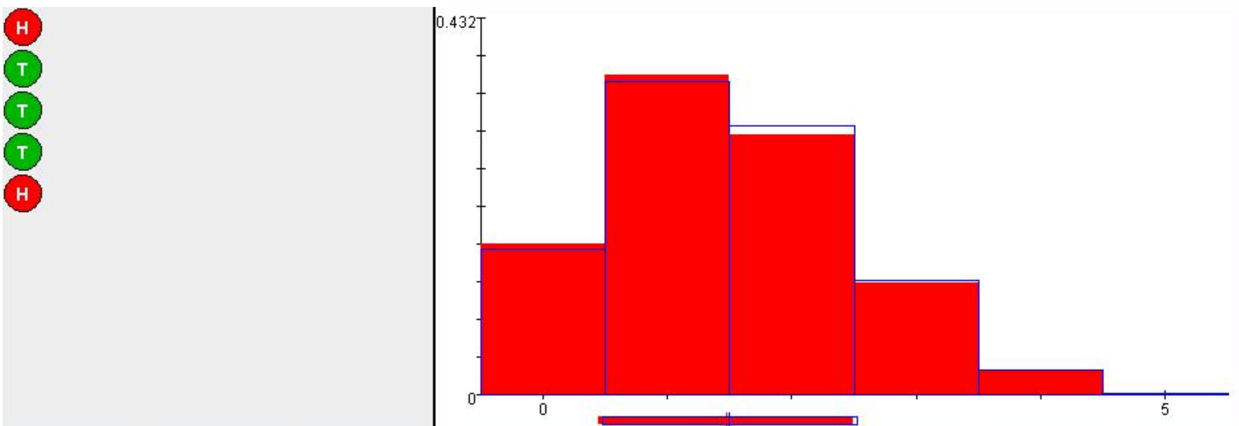


Run	Y	M	U	V
1,000	2	0.4	40	75

Y	Distribution	Data
0	0.16807	0.18
1	0.36015	0.358
2	0.3087	0.3
3	0.1323	0.133
4	0.02835	0.024
5	0.00243	0.005
Mean	1.5	1.478
SD	1.0247	1.03952

5.D (7 POINTS): NOW RUN THE BINOMIAL COIN EXPERIMENT WITH N=5 AND P= 0.3. RUN THE BINOMIAL EXPERIMENT 1,000 TIMES? WHAT IS THE PROPORTION OF OBSERVATIONS THAT HAVE ZERO OR ONE HEAD IN THEM? RECORD THIS VALUE ALSO. HOW CLOSE IS THE PROPORTION VALUE YOU OBTAINED BEFORE TO THIS SAMPLE PROPORTION VALUE? IS THERE A REASON TO EXPECT THAT THESE TWO QUANTITIES (COMING FROM TWO DISTINCT EXPERIMENTS AND TWO DIFFERENT UNDERLYING PROBABILITY MODELS) SHOULD BE SIMILAR? EXPLAIN.

The number of observation with 0 or 1 head are 1000*(.174+.368)=542. This is quite close to my ball and urn experiment of 538. When using replacement, these two experiments indeed follow the same binomial distribution. If I do not use replacement, then the ball and urn experiment follows the hypergeometric distribution. From our qualitative observations earlier, we saw that the variance decreases if we do not use replacement, but the mean remains the same. Therefore, we would still expect our results to be similar.



Run	X	M
1,000	2	0.4

X	Distribution	Data
0	0.16807	0.174
1	0.36015	0.368
2	0.3087	0.299
3	0.1323	0.129
4	0.02835	0.028
5	0.00243	0.002
Mean	1.5	1.475
SD	1.0247	1.02393