Probability & Statistics for EECS: Homework #2 Solution

Problem 1

Given $n \geq 2$ numbers $(a_1, a_2, \ldots, a_n)$ with no repetitions, a bootstrap sample is a sequence $(x_1, x_2, \ldots, x_n)$ formed from the a_j's by sampling with replacement with equal probabilities. Bootstrap samples arise in a widely used statistical method known as the bootstrap. For example, if $n = 2$ and $(a_1, a_2) = (3, 1)$, then the possible bootstrap samples are $(3, 3), (3, 1), (1, 3)$, and $(1, 1)$.

(a) How many possible bootstrap samples are there for $(a_1, \ldots, a_n)$?

(b) How many possible bootstrap samples are there for $(a_1, \ldots, a_n)$, if order does not matter (in the sense that it only matters how many times each a_j was chosen, not the order in which they were chosen)?

(c) One random bootstrap sample is chosen (by sampling from $a_1, \ldots, a_n$ with replacement, as described above). Show that not all unordered bootstrap samples (in the sense of (b)) are equally likely. Find an unordered bootstrap sample $\mathbf{b}_1$ that is as likely as possible, and an unordered bootstrap sample $\mathbf{b}_2$ that is as unlikely as possible. Let p_1 be the probability of getting $\mathbf{b}_1$ and p_2 be the probability of getting $\mathbf{b}_2$ (so p_i is the probability of getting the specific unordered bootstrap sample $\mathbf{b}_i$). What is p_1/p_2? What is the ratio of the probability of getting an unordered bootstrap sample whose probability is p_1 to the probability of getting an unordered sample whose probability is p_2?

Solution

(a) By the multiplication rule, there are n^n possibilities.

(b) By Bose-Einstein, there are $\begin{pmatrix} n+n-1 \\ n \end{pmatrix} = \begin{pmatrix} 2n-1 \\ n \end{pmatrix}$ possibilities.

(c) We can take $\mathbf{b}_1 = [a_1, a_2, \ldots, a_n]$ and $\mathbf{b}_2 = [a_1, a_1, \ldots, a_1]$ (using square brackets to distinguish these unordered lists from the ordered bootstrap samples); these are the extreme cases since the former has $n!$ orders in which it could have been generated, while the latter only has 1 such order. Then $p_1 = n!/n^n, p_2 = 1/n^n$, so $p_1/p_2 = n!$. There is only 1 unordered sample of the form of $\mathbf{b}_1$ but there are n of the form of $\mathbf{b}_2$, so the ratio of the probability of getting a sample whose probability is p_1 to the probability of getting a sample whose probability is p_2 is $(n!/n^n) \,/\, (n/n^n) = (n-1)!$.

Problem 2

If each box of the broad noodle of chili oil flavor contains a coupon, and there are 108 different types of coupons. Given $n \geq 200$, what is the probability that buying n boxes can collect all 108 types of coupons? You need to plot a figure (you do NOT need to submit the code, if used, this time) to show how such probability changes with the increasing value of n. When such probability is no less than 95%, what is the minimum number of n?

Solution

The probability that buying n boxes can collect all m types of coupons can be expressed as:

$$P(n, m) = \frac{m!}{m^n} S(n, m)$$

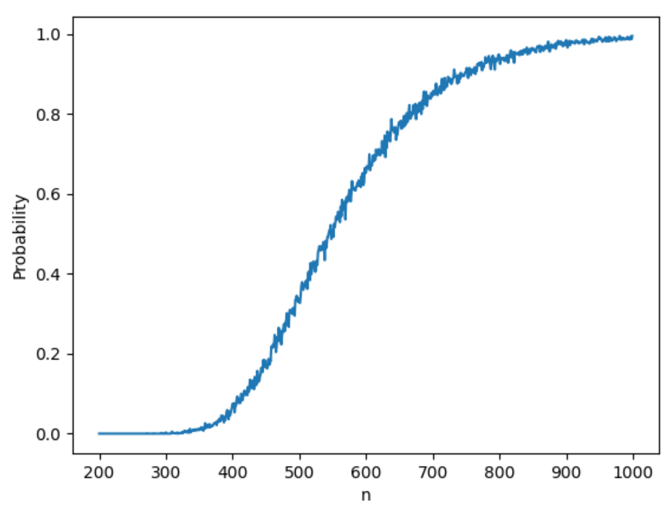
where $S(n, m)$ is the Stirling number of the second kind (count the number of ways to partition a set of n labelled objects into k nonempty unlabeled subsets), and

$$S(n, m) = \frac{1}{m!} \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n$$

In this problem, we have $m = 108$, so the probability that buying n boxes can collect all 108 types of coupons can be expressed as:

$$\begin{aligned} P(108, n) &= \frac{108!}{108^n} S(n, 108) \\ &= \frac{1}{108^n} \sum_{k=0}^{108} (-1)^k \binom{108}{k} (108-k)^n \end{aligned}$$

Then, plot the figure by sampling as below:



By numerical calculation, the minimum number of n to make the probability no less than 95% is $n = 823$.

Problem 3

A batch of one hundred garage kits is inspected by testing four randomly selected ones. If one of the four is defective, the batch is rejected. What is the probability that the batch is accepted if it contains five defectives?

Solution

To choose 4 items from 100 items in one batch, we have $\left(\begin{array}{c} 100 \\ 4 \end{array} \right)$ choice. If the batch is accepted, we should choose 4 items from 95 items which is not defective, and the number of choice is $\left(\begin{array}{c} 95 \\ 4 \end{array} \right)$. So the probability that the batch is accepted is as below:

$$\left(\begin{array}{c} 95 \\ 4 \end{array} \right) / \left(\begin{array}{c} 100 \\ 4 \end{array} \right) \approx 0.812$$

Problem 4

There are three boxes:

a. A box containing two gold coins;

b. A box containing two silver coins;

c. A box containing one gold coin and a silver coin.

After choosing a box randomly and withdrawing one coin randomly, if that happens to be a gold coin, find the probability of the next coin drawn from the same box also being a gold coin.

Solution

Let $G1, G2$ be the event of first and the second coin drawn is a gold coin. Let A, B, C be the event of the box chosen is a, b, c respectively.
So, the probability of the next coin drawn from the same box also being a gold coin can be expressed as below:

$$\begin{aligned}
P(G_2|G_1) &= P(G_2|A, G_1)P(A|G_1) + P(G_2|B, G_1)P(B|G_1) + P(G_2|C, G_1)P(C|G_1) \\
&= 1 \times P(A|G_1) + 0 + 0 \\
&= \frac{P(G_1|A)P(A)}{P(G_1)} \\
&= \frac{P(G_1|A)P(A)}{P(G_1|A)P(A) + P(G_1|B)P(B) + P(G_1|C)P(C)} \\
&= \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 0 + \frac{1}{2} \times \frac{1}{3}} \\
&= \frac{2}{3}
\end{aligned}$$

Problem 5

Mirana is about to play a two-game Starcraft match with an opponent, and wants to find the strategy that maximizes his winning chances. Each game ends with either a win by one of the players, or a draw. If the score is tied at the end of the two games, the match goes into a sudden-death mode, and the players continue to play until the first time one of them wins a game (and the match). Mirana has two playing styles, *i.e.*, timid and bold, and she can choose one of the two at will in each game, no matter what style she chose in previous games. With timid play, she draws with probability $p_d > 0$, and she loses with probability $(1 - p_d)$. With bold play, she wins with probability p_w, and she loses with probability $(1 - p_w)$. Mirana will always play bold during sudden death, but may switch style between games 1 and 2.
Find the probability that Mirana wins the match for each of the following strategies:

1. Play bold in both games 1 and 2.
2. Play timid in both games 1 and 2.
3. Play timid whenever she is ahead in the score, and play bold otherwise.
4. Assume that $p_w < 1/2$, so Mirana is the worse player, regardless of the playing style she adopts. Show that with the strategy in (c) above, and depending on the values of p_w and p_d, Mirana may have a better than a 50-50 chance to win the match. Intuitively, how do you explain this advantage?

Solution

Figure 1.1 provides a sequential description for the three different strategies. Here we assume 1 point for a win, 0 for a loss, and 1/2 point for a draw. In the case of a tied 1-1 score, we go to sudden death in the next game, and Mirana wins the match (probability p_w), or loses the match (probability $1 - p_w$).

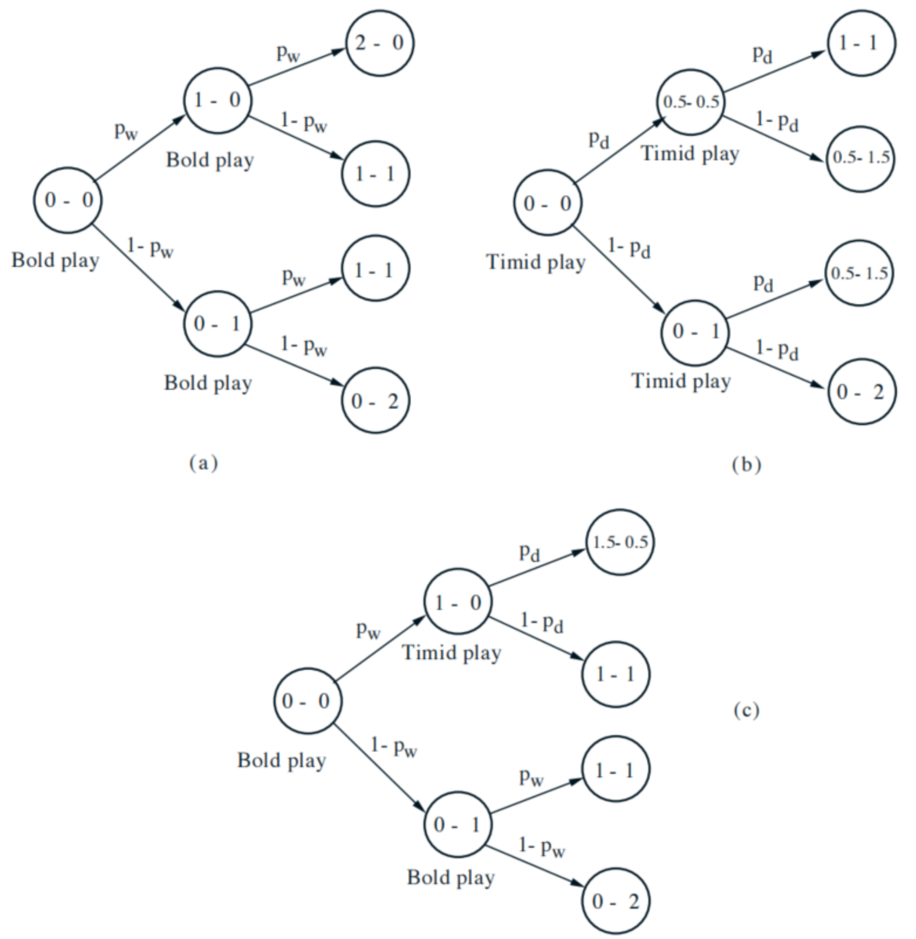


Figure 1.1: Sequential descriptions of the chess match histories under strategies (i), (ii), and (iii).

(1) Using the total probability theorem and the sequential description of Fig. 1.1(a), we have

$$\mathbf{P}(\text{Mirana wins}) = p_w^2 + 2p_w\,(1-p_w)\,p_w.$$

The term p_w^2 corresponds to the win-win outcome, and the term $2p_w\,(1-p_w)\,p_w$ corresponds to the win-lose-win and the lose-win-win outcomes.

(2) Using Fig. 1.1(b), we have

$$\mathbf{P}(\text{Mirana wins}) = p_d^2 p_w$$

corresponding to the draw-draw-win outcome.

Problem 5 continued on next page...

(3) Using Fig. 1.1(c), we have

$$\mathbf{P}(\text{Mirana wins}) = p_w p_d + p_w (1 - p_d) p_w + (1 - p_w) p_w^2$$

The term $p_w p_d$ corresponds to the win-draw outcome, the term $p_w (1 - p_d) p_w$ corresponds to the win-lose-win outcome, and the term $(1 - p_w) p_w^2$ corresponds to lose-winwin outcome.

(b) If $p_w < 1/2$, Mirana has a greater probability of losing rather than winning any one game, regardless of the type of play he uses. Despite this, the probability of winning the match with strategy (3) can be greater than $1/2$, provided that p_w is close enough to $1/2$ and p_d is close enough to 1 . As an example, if $p_w = 0.45$ and $p_d = 0.9$, with strategy (3) we have

$$\mathbf{P}(\text{Mirana wins}) = 0.45 \cdot 0.9 + 0.45^2 \cdot (1 - 0.9) + (1 - 0.45) \cdot 0.45^2 \approx 0.54$$

With strategies (1) and (2), the corresponding probabilities of a win can be calculated to be approximately 0.43 and 0.36 , respectively. What is happening here is that with strategy (3), Mirana is allowed to select a playing style after seeing the result of the first game, while his opponent is not. Thus, by being able to dictate the playing style in each game after receiving partial information about the match's outcome, Mirana gains an advantage.