CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Week 9 Discussion

Conditional Probabilities and Applications

The Birthday Paradox

You are recording the birthday of every person you happen to encounter. Assume for convenience that there are 30 days in each month (and 12 months).

1. **What is the probability that the i^{th} person you encounter has a birthday in March?** We can frame this question in terms of balls and bins. We have 360 bins - each representing a possible birthday. We have n balls, and when we throw a ball it lands in each bin with equal probability $\frac{1}{360}$. Now we can rephrase the question: what's the probability that ball i lands in one of the 30 bins representing March? The ball lands in each of the 30 desired bins with equal probability, so the answer is $\frac{30}{360} = \frac{1}{12}$.

2. **What is the probability that it takes more than n people to find two people who have the same birthday?** Here's the question rephrased in terms of balls and bins: if we throw n balls, what is the probability that no two balls land in the same bin?

   There are two ways to approach this problem. We can first use counting techniques as follows. The first ball can land in any bin. The second can land in any bin other than the bin of the first ball - this occurs with probability $\frac{359}{360}$. Continuing in this way, the n^{th} ball can land in any of the bins not occupied by the first $n-1$ balls, which occurs with probability $\frac{360-(n-1)}{360}$. The probability that it takes more than n people to find two people who have the same birthday is then $\frac{359}{360} \cdots \frac{360-(n-1)}{360}$.

   We can also do this more formally. Let's call the desired event A. Let A_i be the event that ball i lands in a bin which does not contain balls $1,\dots,i-1$. Then $\Pr[A] = \Pr[\cap_{i=1}^{n} A_i] = \Pr[A_1] \times \Pr[A_2|A_1] \times \Pr[A_3|A_1 \cap A_2] \times \cdots \times \Pr[A_n| \cap_{i=1}^{n-1} A_i]$. Observe that $\Pr[A_j| \cap_{i=1}^{j-1} A_i] = \frac{360-(j-1)}{360}$, so $\Pr[A]$ is the same as computed above.

3. **What is the probability that it takes exactly n people to find two people who have the same birthday?** We can again put this question in the balls and bins framework. Like the above question, we are throwing n balls and we have 360 bins. What is the probability that the first $n-1$ balls land in distinct bins and the final ball lands in one of these $n-1$ bins? The probability that the first $n-1$ balls land in distinct bins is given in the previous part - it is $\frac{359}{360} \cdots \frac{360-(n-2)}{360}$. The final ball must land in one of the previous $n-1$ bins, which occurs with probability $\frac{n-1}{360}$. We obtain a final answer of $\frac{359}{360} \cdots \frac{360-(n-2)}{360} \cdot \frac{n-1}{360}$.

4. **Now you're just looking for someone with the same birthday as you. What is the probability that it takes exactly n people for this to occur?** We can again use the same framework: n balls, 360 bins. However, now you pick one bin (which represents your birthday). What's the probability that the first $n-1$ balls do not land in this bin and the n^{th} ball does land in this bin? The probability that the first $n-1$ balls do not land in this bin is $(\frac{359}{360})^{n-1}$ - the balls can land in any of the 359 other

bins. The probability that the final ball lands in the chosen bin is $\frac{1}{360}$. The final probability is then $(\frac{359}{360})^{n-1} \times \frac{1}{360}$.

Let's say you stop recording birthdays after you encounter n people.

1. **What is the probability that exactly 4 of these people were born on March 1?** Out of the n balls, what is the probability that 4 of these balls land in the bin representing March 1? Let's say we specify which balls these should be - we want balls 1,2,3 and 4 to land in the March 1 bin. The probability of this event is $(\frac{1}{360})^4(\frac{359}{360})^{n-4}$. But now we could have specified any choice of 4 balls. There are $\binom{n}{4}$ choices of 4 balls. Once we consider all the choices, the probability that 4 balls land in the bin representing March 1 is $\binom{n}{4}(\frac{1}{360})^4(\frac{359}{360})^{n-4}$.

   We can formalize this argument using the union bound. Let S be the set of all possible choices of 4 balls. For $j \in S$, let A_j be the event that the balls specified by j land in the March 1 bin and all other balls do not land in the March 1 bin. Let A be the event that 4 balls land in the bin representing March 1. Then $\Pr[A] = \Pr[\cup_{j \in S} A_j]$. As we saw above, $\Pr[A_j] = (\frac{1}{360})^4(\frac{359}{360})^{n-4}$. Since $|S| = \binom{n}{4}$, we obtain the same answer as above.

2. **What is the probability that none of these people have birthdays which occur in the months of January, February or March?** Let the event A_i be the event that ball i does not land in any of the bins representing January, February or March. The probability of event A_i is $\frac{9}{12}$. We're trying to compute the probability that none of the n balls land in bins representing January, February or March - so we're trying to compute the probability of $\cap_{i=1}^{n} A_i$. Since these events are independent, we can simply multiply the probabilities, obtaining a final answer of $(\frac{9}{12})^n$.

3. **What is the probability that there exists a person among those n people with a birthday in July?** Let event A_i be the event that ball i lands in a bin representing July. We'd like to compute $\Pr[\cup_{i=1}^{n} A_i]$. Instead, we can compute 1 - $\Pr[\cap_{i=1}^{n} \bar{A}_i]$, where $\bar{A}_i$ is the event that ball i does not land in a bin representing July. This is because the event that there exists a person with a birthday in July is the complement of the event that all n people do not have birthdays in July.

   $\Pr[\bar{A}_i]$ is equal to $\frac{11}{12}$ (person i's birthday can be in any month other than July), and since the events $\bar{A}_i$ are independent, $\Pr[\cap_{i=1}^{n} \bar{A}_i] = (\frac{11}{12})^n$. Therefore, the probability that there exists a person among n with a birthday in July is $1 - (\frac{11}{12})^n$.

Bayesian Inference

The next CS 70 midterm will be cancelled if there is a power outage on campus. Assume the probability of a power outage on any given day is $\frac{1}{100}$. However, one of your GSIs claims to have the ability to predict power outages. When there is actually a power outage, your GSI is correct 90% of the time. When there is not, your GSI is correct only 50% of the time. If your GSI says there will be a power outage on November 5th, what is the probability that he or she is correct?

Let A be the event that the CS 70 midterm is canceled, which is equivalent to the probability of a power outage. Then $\Pr[A] = \frac{1}{100}$. Let B be the event that your GSI predicts that there will be a power outage on November 5th. We know if there is a power outage, your GSI will predict it with probability .9, so $\Pr[B|A] = .9$. We know that if there is not a power outage, your GSI will predict it with probability .5, so $\Pr[B|\bar{A}] = .5$. We would like to calculate $\Pr[A|B]$ - what is the probability that there is a power outage given that your GSI predicts it?

We can use Bayes' Rule - $\Pr[A|B] = \frac{\Pr[B|A]\Pr[A]}{\Pr[B]}$. We already know the values of $\Pr[B|A]$ and $\Pr[A]$, but we need to find $\Pr[B]$. To do this, we use the total probability rule: $\Pr[B] = \Pr[B|A]\Pr[A] + \Pr[B|\bar{A}](1-\Pr[A])$. We can simply plug in values and we obtain that $\Pr[B] = .9 \times .01 + .5 \times .99 = .504$. Now we can go back and plug in values to Bayes' Rule: $\Pr[A|B] = \frac{.9 \times .01}{.504}$ which is approximately .02. Unfortunately, if your GSI says that there will be a power outage on November 5th, it seems very unlikely that there actually will be!