Lecture 12: Applications of Probability

Eric Vigoda
Discrete Mathematics for Computer Science

February 23, 2026

12 Probability Applications

In this lecture we explore several important applications of probability theory that connect directly to computer science. We begin by reviewing conditional probability and Bayes Law, and introduce the Law of Total Probability, using our familiar setting of card examples. We then analyze the classic balls-into-bins model, introducing independence and the Union Bound to prove that hashing achieves logarithmic maximum load with high probability. Finally, we examine Bayesian inference, illustrating how prior beliefs and observed data combine to determine posterior probabilities in both coin-tossing and medical testing scenarios.

12.1 Conditional Probability and Bayes Law Review

Recall, the definition of conditional probability from last lecture. For events A and B where $\Pr[B] > 0$ then we have the following:

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

The way to remember the above equation is that when we condition on the event B then we need to renormalize the state space from Ω down to B, and then to ensure that the subsequent probabilities sum to 1, then for each $\omega \in B$ we need to rescale $Pr[\omega] \to \Pr[\omega]/\Pr[B]$.

We then obtain **Bayes Law**:

$$\Pr[A \mid B] = \frac{\Pr[B \mid A]\Pr[A]}{\Pr[B]}.$$

This is useful when $\Pr[B \mid A]$ is easier to compute than $\Pr[A \mid B]$; we'll see several examples later in the lecture.

To see why Bayes Law is true, notice that from the definition of conditional probability we have the following which is known as the **chain rule**:

$$\Pr[A \cap B] = \Pr[A \mid B]\Pr[B].$$

The chain rule gives us a simple way to "atomize" compound events.

Example 1: Suppose we draw 2 cards at random from a deck of 52 cards, what's the probability that both cards are Aces?

Let A denote the event that both cards are Aces. Let A_1 denote the event that the 1st card is an Ace, and let A_2 denote the event that the 2nd card is an Ace. Then,

$$\Pr[A] = \Pr[A_1 \cap A_2] = \Pr[A_2 \mid A_1]\Pr[A_1] = \frac{3}{51} \times \frac{4}{52} = \frac{12}{51 \times 52}.$$

We can easily verify this using a direct calculation:

$$\Pr[A] = \frac{\binom{4}{2}}{\binom{52}{2}} = \frac{4 \times 3/2}{52 \times 51/2} = \frac{12}{51 \times 52}.$$

Applying the chain rule, we obtain what's called the **Law of Total Probability** (these terms are not important but the hopefully intuitive equations are important and often very useful):

$$\begin{aligned}\Pr[A] &= \Pr[A \mid B] \Pr[B] + \Pr[A \mid \overline{B}] \Pr[\overline{B}] \\ &= \Pr[A \cap B] + \Pr[A \cap \overline{B}],\end{aligned}$$

which is saying that for the event A to occur then either B also occurs or $\overline{B}$ occurs, and in each case we can apply the chain rule to compute the probability of the intersection either of $A \cap B$ or $A \cap \overline{B}$.

Let's look at an example application of the Law of Total Probability.

Example 2: Once again, draw 2 cards at random from a deck of 52 cards. What's the probability that at least one of the cards is an Ace?

Let A denote the event that at least one card is an Ace, and let B denote the event that the 1st card is an Ace. Then,

$$\Pr[A] = \Pr[A \mid B] \Pr[B] + \Pr[A \mid \overline{B}] \Pr[\overline{B}] = 1 \times \frac{4}{52} + \frac{4}{51} \times \frac{48}{52} = \frac{4 \times 51 + 4 \times 48}{51 \times 52} = \frac{4 \times 99}{51 \times 52}.$$

Let's compare this to a direct calculation using complement counting:

$$\Pr[A] = 1 - \Pr[\overline{A}] = 1 - \frac{\binom{48}{2}}{\binom{52}{2}} = 1 - \frac{48 \times 47/2}{52 \times 51/2} = 1 - \frac{48 \times 47}{52 \times 51} = \frac{52 \times 51 - 48 \times 47}{51 \times 52} = \frac{4 \times 99}{51 \times 52},$$

which exactly matches the expression we obtained from applying the Law of Total Probability. Note in the above calculation, when simplifying the numerator we used the following algebra:

$$52 \times 51 - 48 \times 47 = 4(13 \times 51 - 12 \times 47) = 4(663 - 564) = 4 \times 99.$$

Now let's look at a similar application of Bayes Law.

Example 3: Draw 2 cards at random from a deck of 52 cards. What's the probability that the first card was an Ace given that at least one of the cards is an Ace.

Let A denote the event that at least one card is an Ace, and let B denote the event that the 1st card is an Ace. Now our goal is to compute $\Pr[B \mid A]$. Notice that $\Pr[A \mid B]$ is trivial since $B \subset A$ and thus $\Pr[A \mid B] = 1$. Hence, let's apply Bayes Law to simplify our calculations:

$$\Pr[B \mid A] = \frac{\Pr[A \mid B] \Pr[B]}{\Pr[A]} = \frac{1 \times \Pr[B]}{\Pr[A]} = \frac{4/52}{1 - \binom{48}{2}/\binom{52}{2}} = \frac{4}{52} \cdot \frac{52 \cdot 51}{52 \cdot 51 - 48 \cdot 47} = \frac{51}{99} = \frac{17}{33}.$$

12.2 Hashing Idealized: Balls into Bins

Let's look at another setting: throwing balls into bins. This will be useful for studying the performance of hashing, as we'll see extensively in CS 130A.

We begin with a brief description of the application to hashing; if you're unfamilar with hashing (or uninterested) then you can safely skip to the below example. Suppose we have a huge universe of possible keys, which we denote as U. We want to compactly maintain a subset $S \subset U$. To do this we use a hash function and a hash table. So we have a hash table $H[1 \dots n]$, and we have a hash function $h : U \to H$.

We're thinking of chain hashing where each $x \in S$ is stored in a linked list at position $h(x)$ of table H. To minimize the query/look-up time, we want to minimize the size of these linked lists. We'll assume our hash function h is random and each choice is independent of the other; this means that for every $x \in U$, $h(x)$ is uniformly chosen from $\{1, \dots, n\}$, and that the choice of $h(x)$ is not dependent on $h(y)$ for any $y \neq x$ (we will formalize the notion of independence in a moment). If the subset S has size $n = |S|$ then we can view each element $x \in S$ as a ball, and the indices in the table H as bins. Hence, our goal is to upper bound the maximum number of balls in any bin, assuming each ball is placed in a random bin, independent of the other balls.

12.2.1 Independent Events

Two events A and B are said to be **independent events** if

$$\Pr[A \cap B] = \Pr[A] \times \Pr[B].$$

This implies that the conditional probability of A given B is equal to A (without any conditioning on B), and hence knowledge of B gives no information about whether A occurs or not:

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\Pr[A] \times \Pr[B]}{\Pr[B]} = \Pr[A].$$

12.3 Balls into Bins

Example 4: We have n identical balls and n bins labeled $B_1, \dots, B_n$. Each ball is placed into a bin chosen uniformly at random from the n bins, and this choice is independent of the placement of the other balls. For $1 \le i \le n$, let $L(i)$ denote the load of bin B_i which is defined as the number of balls placed in bin B_i. What is the probability that the maximum load is $\le 10 \log n$? Here and throughout this section, log denotes base 2.

We will show that the maximum load is $> 10 \log n$ with very small probability, and hence in our chain hashing example the query or lookup-time is $O(\log n)$ with high probability.

12.4 Analyzing Max-Load for Balls into Bins

For each bin B_i, define the event

$$A_i = \{L(i) \ge 10 \log n\},$$

that is, A_i is the event that bin B_i receives at least $10 \log n$ balls.

Let M denote the maximum load:

$$M = \max_{1 \le i \le n} L(i).$$

Observe that the event $\{M \ge 10 \log n\}$ occurs if and only if at least one of the events A_i occurs. In other words,

$$\{M \ge 10 \log n\} = \bigcup_{i=1}^{n} A_i.$$

Now we apply the **Union Bound**, which states that for any events $E_1, \dots, E_n$,

$$\Pr\left[\bigcup_{i=1}^{n} E_i\right] \le \sum_{i=1}^{n} \Pr[E_i].$$

The Union Bound can be viewed as a very coarse version of inclusion–exclusion, where we simply ignore the overlap between events. Or you can directly observe that for subsets $S_1, \dots, S_n \subset \Omega$:

$$|S_1 \cup \cdots \cup S_n| \le |S_1| + \cdots + |S_n|,$$

because elements in overlaps are counted multiple times on the right-hand side.

Applying the Union Bound to the events $A_1, \dots, A_n$, we obtain

$$\Pr[M \ge 10 \log n] = \Pr\left[\bigcup_{i=1}^{n} A_i\right] \le \sum_{i=1}^{n} \Pr[A_i].$$

Since all bins are symmetric,

$$\Pr[A_i] = \Pr[L(1) \ge 10 \log n] \quad \text{for every } i.$$

Hence,

$$\Pr[M \ge 10 \log n] \le n \Pr[L(1) \ge 10 \log n].$$

We now bound $\Pr[L(1) \geq k]$ for a general integer k, and then we set $k = 10 \log n$ to derive our desired bound. For any $k \geq 1$, the probability the load at bin B_1 is exactly k is equal to the probability that exactly k balls choose bin B_1 and the other $n-k$ balls choose one of the other $n-1$ bins, thus:

$$\Pr[L(1) = k] = \binom{n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k}.$$

This distribution for $L(1)$ is known as the Binomial distribution and is denoted as

$$L(1) \sim \mathrm{Bin}(n, 1/n).$$

Similarly, to obtain load $\geq k$, then there must exist some subset of k balls that all chose bin B_1 (and the other $n-k$ balls can go to any of the n bins including B_1), and thus:

$$\{L(1) \geq k\} = \bigcup_{\substack{S \subseteq [n]: \\ |S| = k}} \{\text{all balls in } S \text{ choose } B_1\},$$

where $[n] = \{1, \ldots, n\}$. Applying the Union Bound,

$$\Pr[L(1) \geq k] \leq \binom{n}{k} \left(\frac{1}{n}\right)^k.$$

Using the binomial bound from HW 6,

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k,$$

we obtain

$$\Pr[L(1) \geq k] \leq \binom{n}{k} \left(\frac{1}{n}\right)^k \leq \left(\frac{en}{k}\right)^k \left(\frac{1}{n}\right)^k = \left(\frac{e}{k}\right)^k.$$

Setting $k = 10 \log n$, we conclude that

$$\Pr[L(1) \geq 10 \log n] \leq \left(\frac{e}{10 \log n}\right)^{10 \log n} \leq \left(\frac{1}{2}\right)^{10 \log n} = \frac{1}{n^{10}},$$

where we used the fact that $10 \log n > 2e$ which is true for all $n \geq 2$ (with $\log n$ being base 2).

Therefore,

$$\Pr[M \geq 10 \log n] \leq n \times \frac{1}{n^{10}} = \frac{1}{n^9}.$$

Hence, with high probability, namely, with probability $\geq 1 - n^{-9}$, then

$$M < 10 \log n.$$

In the context of chain hashing with a random hash function, this shows that the maximum linked list size is $O(\log n)$ with high probability.

12.5 Bayesian Inference: Priors and Posteriors

Now let's return to a more sophisticated and powerful application of Bayes Law to Bayesian inference.

Bayes Law gives us a powerful way to update our beliefs when new information becomes available. The prior probability $\Pr[S]$ represents our belief before seeing the new evidence. The conditional probability $\Pr[S|T]$ represents our updated belief after observing T. This updated probability is often called the **posterior probability**.

In many real situations, we *cannot* compute a posterior probability without first specifying a prior belief. This is one of the main takeaways of Bayesian inference: the data updates your beliefs, but it does not replace them.

12.5.1 Bayesian Inference: Coin Tossing

Suppose a coin is chosen in the following way. With probability π it is a **biased coin** with $\Pr[H] = 3/4$ (and hence $\Pr[T] = 1/4$), and with probability $1 - \pi$ it is a **fair coin** with $\Pr[H] = 1/2$.

We do not know which coin we were given. We toss it 10 times and observe exactly 8 heads.

Let S be the event "the coin is biased" and let F be the event "the coin is fair." Let D be the event "in 10 tosses, we see exactly 8 heads."

Our goal is to compute the posterior probability $\Pr[S|D]$.

Likelihoods. By the binomial formula, we have:

$$\Pr[D|S] = \binom{10}{8}\left(\frac{3}{4}\right)^8\left(\frac{1}{4}\right)^2 \approx 0.2816 \qquad \text{and} \qquad \Pr[D|F] = \binom{10}{8}\left(\frac{1}{2}\right)^{10} \approx 0.0439.$$

At this point, we could use the likelihoods to conclude that it is more likely that the biased coin was used. Or we sample proportional to the likelihoods to say that with probability $.2816/(.2816 + .0439) \approx .865$ the biased coin was used and with probability $\approx .135$ the fair coin was used. But these estimates are based on the probability of seeing the data D under each of these settings.

We want to look at the probability of each of these settings (or models) given the data D we observed; these are called the **posterior probabilities**. To obtain the posterior probabilities we apply Bayes Law:

$$\Pr[S \mid D] = \frac{\Pr[D \mid S]\Pr[S]}{\Pr[D]}$$
$$\Pr[F \mid D] = \frac{\Pr[D \mid F]\Pr[F]}{\Pr[D]}$$

The first terms in the numerator, namely $\Pr[D \mid S]$ and $\Pr[D \mid F]$, are the likelihoods which we just computed.

What does the second term in the numerator mean? In particular, what is $\Pr[S]$ and $\Pr[F]$? These are called the **priors**. These represent your belief or confidence, **prior to seeing the experimental results**, in whether a biased or fair coin was used. What about the denominator? This is simply a normalizing factor so that the quantities sum up to 1, since either S occurs or F occurs. To obtain this normalizing factor, we set $\Pr[D] = \Pr[D \mid S]\Pr[S] + \Pr[D \mid F]\Pr[F]$, which is the sum of the numerators.

Let's look at 2 different priors.

Recall, $\Pr[S] = \pi$ and $\Pr[F] = 1 - \pi$; these are our priors. Then applying Bayes Law as done above we obtain the following:

$$\Pr[S \mid D] = \frac{\Pr[D \mid S]\Pr[S]}{\Pr[D]} = \frac{\Pr[D \mid S]\pi}{\Pr[D|S]\pi + \Pr[D \mid F](1-\pi)}.$$

A convenient way to simplify this calculation is to consider the likelihood ratio:

$$L \;=\; \frac{\Pr[D|S]}{\Pr[D|F]} = \frac{\binom{10}{8}\left(\frac{3}{4}\right)^8\left(\frac{1}{4}\right)^2}{\binom{10}{8}\left(\frac{1}{2}\right)^{10}} = \frac{\left(\frac{3}{4}\right)^8\left(\frac{1}{4}\right)^2}{\left(\frac{1}{2}\right)^{10}} \approx 6.4072.$$

Then

$$\Pr[S|D] = \frac{L\pi}{L\pi + (1-\pi)}.$$

Two different priors give two different posteriors.

1. **Prior $\pi = \frac{1}{2}$.**

$$\Pr[S|D] = \frac{L/2}{L/2 + 1/2} = \frac{L}{L+1} \approx \frac{6.4072}{7.4072} \approx 0.8650.$$

   So with a 50-50 prior, seeing 8 heads out of 10 makes it quite likely the coin is biased. This prior is called a uniform prior (for obvious reasons). And notice that the posterior probabilities with a uniform prior are exactly the same as the probabilities obtained from using the likelihoods; if we use a uniform prior, then we don't have any prior knowledge so the posterior probabilities should be the same as the likelihoods. To be precise, they are not "the same" as likelihoods; they are normalized likelihoods.

2. **Prior $\pi = 0.01$.**
$$\Pr[S|D] = \frac{0.01\,L}{0.01\,L + 0.99} \approx \frac{0.06407}{1.05407} \approx 0.0608.$$
Now the *same data* leads to a very different conclusion: even after seeing 8 heads out of 10, the posterior probability the coin is biased is only about 6%.

Discussion. The data contributes a multiplicative factor (here, the likelihood ratio $L \approx 6.41$), but if your prior belief π is extremely small, then a modest amount of evidence might not be enough to overcome it. This is exactly why, in Bayesian inference, specifying the prior is not optional: the posterior depends on it.

12.5.2 Bayesian Inference: Medical Testing

Now let's look at a more realistic (and more surprising) application.

Suppose there is a disease that affects a fraction of the population. We have a specific individual who underwent a medical test for the disease, and we want to determine the probability the individual has the disease based on their test results.

Let S denote the event "the person has the disease," and let $\overline{S}$ denote the event "the person does not have the disease."

Let T denote the event "the test result is positive."

Prior probabilities.
Suppose the disease affects 1% of the population. Then:
$$\Pr[S] = 0.01, \qquad \Pr[\overline{S}] = 0.99.$$
These are the **priors**.

Likelihoods.
Suppose the medical test is 99% accurate:
$$\Pr[T|S] = 0.99, \qquad \Pr[T|\overline{S}] = 0.01.$$
This means the following:

- The false negative rate, which is $\Pr[\overline{T}|S]$, is 1%,
- The false positive rate, which is $\Pr[T|\overline{S}]$, is 1%.

Posterior probability.
If a person tests positive, what is the probability they actually have the disease? We compute the posterior probability $\Pr[S|T]$ using Bayes Law:
$$\Pr[S|T] = \frac{\Pr[T|S]\Pr[S]}{\Pr[T|S]\Pr[S] + \Pr[T|\overline{S}]\Pr[\overline{S}]}.$$
Substituting:
$$= \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.01 \cdot 0.99}.$$
Observe that the numerator and denominator terms are equal:
$$= \frac{0.0099}{0.0099 + 0.0099} = \frac{0.0099}{0.0198} = \frac{1}{2}.$$
Thus,
$$\Pr[S|T] = 0.5.$$

This is often surprising. Even though the test is 99% accurate, a positive test result only means there is a 50% chance the person actually has the disease. The reason is that the disease is rare, so false positives occur almost as often as true positives.

Now suppose instead that 10% of the population has the disease:

$$\Pr[S] = 0.10, \qquad \Pr[\overline{S}] = 0.90.$$

The test accuracy remains the same.
Applying Bayes Law again:

$$\Pr[S|T] = \frac{0.99 \cdot 0.10}{0.99 \cdot 0.10 + 0.01 \cdot 0.90}.$$

Simplifying:

$$= \frac{0.099}{0.099 + 0.009} = \frac{0.099}{0.108} \approx 0.917.$$

Now the posterior probability is about 91.7%.

The only thing that changed was the **prior probability** $\Pr[S]$. The test accuracy remained the same. This demonstrates a key lesson of Bayesian inference:

Evidence cannot be interpreted without taking prior probabilities into account.

Varying the test sensitivity. Variation: Asymmetric Error Rates
Now suppose the test has a 10% false negative rate and a 1% false positive rate. Thus:

$$\Pr[T|S] = 0.90, \qquad \Pr[T|\overline{S}] = 0.01.$$

Case 1: Disease prevalence 1%.
Suppose

$$\Pr[S] = 0.01, \qquad \Pr[\overline{S}] = 0.99.$$

Applying Bayes Law:

$$\Pr[S|T] = \frac{\Pr[T|S]\Pr[S]}{\Pr[T|S]\Pr[S] + \Pr[T|\overline{S}]\Pr[\overline{S}]}.$$

Substituting:

$$= \frac{0.90 \cdot 0.01}{0.90 \cdot 0.01 + 0.01 \cdot 0.99} = \frac{0.009}{0.009 + 0.0099} = \frac{0.009}{0.0189} \approx 0.476.$$

Even with a highly accurate test, the posterior probability is still below 50% because the disease is rare and false positives remain significant.

Case 2: Disease prevalence 10%.
Now suppose

$$\Pr[S] = 0.10, \qquad \Pr[\overline{S}] = 0.90.$$

Applying Bayes Law again:

$$\Pr[S|T] = \frac{0.90 \cdot 0.10}{0.90 \cdot 0.10 + 0.01 \cdot 0.90} = \frac{0.09}{0.09 + 0.009} = \frac{0.09}{0.099} \approx 0.909.$$

Now the posterior probability exceeds 90%.

This again illustrates the central lesson of Bayesian inference:

Posterior probabilities depend critically on prior probabilities.