CS 2800 Review

Pradhyum Rajasekar

14 Nov 2025 – present

1 Probability

Consider a randomized experiment. Then:

- the set of **outcomes** Ω refers to all possible results of the experiment.
- an **event** A is a subset of the outcome space. The set of all events $\mathcal{F}$ must obey these conditions:
  1. $\Omega \in \mathcal{F}$
  2. $A \in \mathcal{F} \implies (\Omega \backslash A) \in \mathcal{F}$. This condition is known as being closed under complements. $\Omega \backslash A$ is also notated as $\overline{A}$.
  3. $A_1 \in \mathcal{F} \wedge A_2 \in \mathcal{F} \implies (A_1 \cup A_2) \in \mathcal{F}$ This condition is known as being closed under countable unions since this union property can apply to any countable number of events.
- A **probability function** $P : \mathcal{F} \to \mathbb{R}^+$ defines the probability of an event A. It follows two basic conditions:
  1. Total Probability: $P(\Omega) = 1$
  2. Countable Additivity: $P(\bigcup A_i) = \sum P(A_i)$ where the A_i are pairwise disjoint and the set of A_i is countable.

  There are 5 additional properties that follow from this:

  1. Null Probability: $P(\emptyset) = 0$
  2. Monotonicity: $A \subseteq B \implies P(A) \leq P(B)$
  3. Boundedness: $0 \leq P(A) \leq 1$
  4. Complement probability: $P(\overline{A}) = 1 - P(A)$
  5. Union Bound: $P(A \cup B) \leq P(A) + P(B)$
- A **probability space** is defined as the triple $(\Omega, \mathcal{F}, P)$.

To illustrate the notation above, consider an experiment where $P(\{\omega\})$, or simply $P(\omega)$ as a shorthand, is uniform for every $\omega \in \Omega$: a situation like a dice roll or a random card from a deck. Then, for any event A (like the card number is even), its probability is $P(A) = \frac{|A|}{|\Omega|}$.

Other terms related to probability:

- **Conditioning**: The probability of an event A given an event B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- **Independence** is defined as $P(A|B) = P(A)$, or, equivalently, $P(A \cap B) = P(A)P(B)$.

- **Mutual Independence**: every subset S of the set of events A_i must follow:

$$P\left(\bigcup_{A_i \in S} A_i\right) = \prod_{A_i \in S} P(A_i)$$

- **Pairwise Independence**: where every pair of sets in some set of events are independent

Again, consider a randomized experiment, where we are only concerned about a particular feature of the experiment. Example: the number of heads when a coin is flipped 100 times. Then:

- A **random variable** $X : \Omega \to \mathbb{R}$ encodes every possible result as a real number. In the example, this function could be counting the number of heads.
- The **support** R_X of a random variable X is its range. A discrete RV has a countable support.
- The **probability mass function (pmf)** $f_X : R_X \to [0,1]$ gives the probability that an event occurs. The **cumulative distribution function (cdf)** $F_X : \mathbb{R} \to [0,1]$ gives the probabilitiy that the random variable is at most the input to the cdf.

There are a common set of kinds of random variables that can be used to solve different problems. This class covers **discrete uniform**, **Bernoulli**, **geometric**, and **binomial** random variables:

Notation	Support	pmf	Notes
$\text{Unif}\{a,b\}$	$\{u \in \mathbb{Z}^+ \mid a \leq u \leq b\}$	$f_U(r) = \frac{1}{b-a+1}$	Equal probability between a and b, inclusive
$\text{Bern}(p)$	$\{0,1\}$	$f_U(0) = 1-p,\ f_U(1) = p$	Boolean that is true with probability p
$\text{Binom}(n,p)$	$\{0,1,\ldots,n\}$	$f_U(r) = \binom{n}{r}p^r(1-p)^{n-r}$	n Bernoulli variables, r being true
$\text{Geom}(p)$	$\mathbb{N}^+$	$f_U(r) = (1-p)^{r-1} \cdot p$	Repeat Bernoulli variable until true

The notation $\mathbb{I}(A)$ where A is an event refers to an **indicator variable**, which is 1 when the event occurs and 0 otherwise; it is equivalent to $\text{Bern}(P(A))$.

Here are some aggregate properties that can be calculated about these variables:

- The **expectated value** of a random variable is the weighted average of values in its support weighted by their probabilities:
$$\mathbb{E}[X] = \sum_{r \in R_X} r \cdot f_X(r)$$
- The **variance** of a random variable is the expected square of the deviation of values from the expected value:
$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$
- The **standard deviation** is the square root of the variance.
- The **law of large numbers** states that, as the number of repeated trials of a random variable increases, the mean of the results approaches the expected value.

Distribution	Expected Value	Variance
Uniform	$\frac{a+b}{2}$	
Bernoulli	p	$p(1-p)$
Binomial	np	$np(1-p)$
Geometric	$\frac{1}{p}$	

When dealing with multiple random variables, it is often useful to use the following definitions and tools:

- A **joint distribution** is a description of the probability that multiple random variables attain certain values.
  - The pmf is a function $f_{X,Y}(r_1, r_2)$ which gives the probability that both $X = r_1$ and $Y = r_2$.
  - The cdf is a function $F_{X,Y}(r_1, r_2)$ which gives the probability that both $X \leq r_1$ and $Y \leq r_2$
- The **marginal distribution** of some random variable in a joint distribution refers to its normal distribution.
- A **conditional distribution** is the distribution given some result of some random variable; it is what it sounds like.
- Two random variables are independent if $f_{X,Y}(r_1, r_2) = f_X(r_1) \cdot f_Y(r_2)$
- Given two independent random variables X and Y, then

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

- The **covariance** of two random variables is defined as

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

  Note that $\text{Var}(X) = \text{Cov}(X, X)$

- The **correlation** is the covariance divided by the standard deviation of the two relevant variables:

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

1.1 Laws to Know

Lemma 1. *Law of Total Probability:*
Let the A_i be a partition of Ω, and let B be some event. Then,

$$P(B) = \sum_i P(A_i)\, P(B|A_i)$$

Lemma 2. *Bayes' Rule:*
Given an event A and B:

$$P(B|A) = \frac{P(A|B)\, P(B)}{P(A)}$$

The Law of Total Probability can be used to calculate the denominator of the RHS in Bayes' Rule.

Lemma 3. *Law of the Unconscious Statistician (LOTUS):*

$$\mathbb{E}[g(x)] = \sum_{r \in R_X} g(r) \cdot f_X(r)$$

Lemma 4.

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Proof.

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2 - 2\mathbb{E}[X] \cdot X + \mathbb{E}[X]^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X] \cdot \mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

□

Lemma 5. *If there are n mutually independent random variables $B_i \sim Bern(p)$, then $B = B_1 + B_2 + \ldots + B_n \sim Binom(n, p)$*

Theorem 1. ***Linearity of Expectation**:*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Theorem 2. ***Something of Variance**: for two* independent *random variables X and Y:*

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$$

Markov's Inequality gives a very loose bound using only the expected value.

Theorem 3. ***Markov's Inequality**: for a nonnegative random variable X,*

$$P(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \sum r \cdot f_X(r) \\
&= \sum_{r<\alpha} r \cdot f_X(r) + \sum_{r\geq\alpha} r \cdot f_X(r) \\
&\geq \sum_{r<\alpha} 0 \cdot f_X(r) + \sum_{r\geq\alpha} \alpha \cdot f_X(r) \\
&= \alpha \cdot P(X \geq \alpha)
\end{aligned}$$

□

Chebyshev's Inequality gives a tigher bound using the variance, giving the likelihood that the random variable is more than β away from the mean.

Theorem 4. ***Chebyshev's Inequality**: given a random variable X, for any β*

$$P(|X - \mathbb{E}[X]| \geq \beta) \leq \frac{\mathrm{Var}(X)}{\beta^2}$$

Proof.

$$\begin{aligned}
P(|X - \mathbb{E}[X]| \geq \beta) &= P((X - \mathbb{E}[X])^2 \geq \beta^2) \\
&\leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\beta^2} \\
&= \frac{\mathrm{Var}(X)}{\beta^2}
\end{aligned}$$

□

2 Exercises

To-do:

- 24.7 (220)
- 25.9 (229)
- 26.10 (239)
- 27.8 (247)
- 28.9 (254)
- 29.6 (261)
- 30.8 (269)
- 31.4 (276)

2.1 Practice Problems

- *Prove that, if A is countable and B is finite, then $A \times B$ is countable.*

  Proof. If A is countable, then there exists some injection $f : A \to \mathbb{N}$; similarly, since B is finite, there exists some bijection $g : B \to [n]$ where n is the size of B. Consider the function $h : A \times B \to \mathbb{N}$:

  $$h(a, b) = 2^{f(a)}3^{g(b)}$$

  Since f and g are both injections, every unique a and b input into the function h will result in unique values of $f(a)$ and $g(b)$; therefore, since every unique input into h results in a unique number since no two numbers can be the same with a differing prime factorization, h is an injection, and therefore $A \times B \leq \aleph_0$. $\square$

- *Suppose that $f \circ f = f$, $f : X \to X$. Prove that f either injective or surjective implies that $f = \mathrm{id}_X$*

  Proof. First, I shall prove that f being injective implies that f is the identity function. Note that $f(x) = f(f(x))$. Being injective implies that there can only be one x which gives the value $f(x)$; applied to this equation, this means that the inputs to the functions must be the same:

  $$x = f(x),$$

  proving the statement.

  Next, I shall prove that f being surjective implies that f is the identity function. Since f is surjective, for any a, there is some b such that $f(b) = a$ and some c such that $f(c) = b$. Therefore, $f(f(c)) = f(c) = a = b$, and since a and b have to be the same, $f(a) = a$. $\square$

- *Alice and Bob play the following game: Bob gets to pick any 10 distinct integers from 1 to 40. Alice has to find two different sets of three numbers that have the same sum. Prove that Alice always wins.*

  Proof. Alice has $\binom{10}{3} = 120$ different sets she can choose from. The minimum possible sum is 6 and the maximum possible sum is 117, leaving an upper bound of 111 possible sums. Since there are more sets than sums, by the pigeonhole principle, at least one sum is associated with more than one set; if Alice picks two of the multiple sets with this sum, she wins. $\square$

- *There are 7 balls, each of which is either red, blue, green, or white. Prove that there are 2 distinct subsets of 3 balls that have the same colors.*

  Proof. There are $\binom{7}{3} = 35$ different subsets of 3 balls. To count the number of possible colorings of balls:

  $$\binom{4}{3} + \binom{4}{2}(3) + \binom{4}{1} = 26$$

  The pigeonhole principle can be applied with the subsets representing the pigeons and the colorings representing the holes: there has to be at least one coloring that is mapped to by more than one subset of balls. Any two subsets can be selected from the set of subsets of size 3 that map to that particular coloring. $\square$

- *Find the number of 6 digit numbers that satisfy:*

  1. *The first or last digit is even*
     If the first digit is even, then it has to be in the set $2, 4, 6, 8$; this leads to $4 \cdot 10^5$ numbers. Similarly, if the last digit is even, then it has to be in the set $0, 2, 4, 6, 8$, leading to $10^5 \cdot 5$. To apply the inclusion-exclusion principle, the number of numbers where both the first and last digits are even is $4 \cdot 10^4 \cdot 5$. The answer is therefore

     $$4 \cdot 10^5 + 5 \cdot 10^5 - 2 \cdot 10^5 = 7 \cdot 10^5$$

2. *The sum of the digits is 10* This can be solved using stars-and-bars: since the first digit has to be 1, there are 9 stars and, since there are 6 digits, 5 bars:

$$\binom{9+5-1}{5} = 1287$$

▶ *Give a combinatorial proof of this identity:*

$$\binom{n}{k} - \binom{n-2}{k} = \binom{n-1}{k-1} + \binom{n-2}{k-1}$$

Proof. Consider the set $S = \{U \subseteq [n] \mid 1 \in U \vee 2 \in U, |U| = k\}$. I will demonstrate two different ways of counting the number of elements in this set. The first method is counting the number of subsets of $[n]$ of size k that do not include 0 or 1 and subtract that number from the total number of subsets of $[n]$ of size k; for the first quantity, since there are 2 excluded elements, the total number of elements that can be selected is $\binom{n-2}{k}$, while the total number of subsets is $\binom{n}{k}$. This gives the LHS as a valid count of the number of elements in S.

Another way of counting the elements of S is by selecting all subsets that include 0, and adding that to the number of subsets that include 1 but not 0; this method does not double count. Since the former quantity can be calculated by fixing one element of the subset, giving $\binom{n-1}{k-1}$, and the latter quantity can be calculated by fixing one element of the subset and further reducing the number of elements by 1 to preclude 0, giving $\binom{n-2}{k-1}$, their sum gives the RHS. □

▶ *Give a combinatorial proof of this identity:*

$$\binom{n}{m}\binom{m}{k} = \binom{n}{k}\binom{n-k}{m-k}$$

Proof. Consider the set $S = \{(U \subseteq V, V \subseteq [n]) \mid |U| = k, |V| = m\}$. This can be counted by selecting all subsets V of $[n]$, which is $\binom{n}{m}$, then selecting all subsets U of V, which is $\binom{m}{k}$, then combining the two numbers using the product rule, yielding the LHS.

Alternatively, a subset U of elements of size k can be selected from $[n]$, and, from the remaining elements not in U, $m-k$ elements can be selected to form another set V' such that $V = V' \cup U$ is of size m and U is a subset of V; since there are $\binom{n}{k}$ to perform the first step and $\binom{n-k}{m-k}$ ways to perform the second step, and they are both independent, they can be combined using the product rule, yielding the RHS. □

▶ *Let I_1 and I_2 be 2 different permutations of $[n]$. What is the expected number of numbers for which it appears in I_1 before I_2?*

Let X_i be the event that i appears in I_1 before I_2. There is a $\frac{1}{n}$ probability that i appears in the same place; since the probability that i appears in I_1 before I_2 and the other way around are the same, the probability that i appears before in I_1 is $\frac{n-1}{2n}$. By linearity of expectation, since there are n X_i, the expected number is $\frac{n-1}{2}$.

▶ *A hat-check staff randomly returns n hats to n people. Let S_n be the number of people who get their own hat back; what is $\mathbb{E}[S_n^2]$?*

Let X_i be an indicator variable for the ith person getting their hat back. Then, since the probability of that happening is $\frac{1}{n}$, $\mathbb{E}[S_n] = 1$.

Now, calculate $\mathbb{E}[X_i X_j]$; this is simply equal to the probability that both the ith and jth person get their hats back. Since, once the ith person gets the hat, the probability of the jth person getting their hat back is increased due to there being one more hat, $\mathbb{E}[X_i X_j] = \frac{1}{n(n-1)}$.

Now, the actual algebra can commence:

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \mathbb{E}[(X_1 + X_2 + \ldots + X_n)^2] \\
&= \mathbb{E}[X_1^2 + X_2^2 + \ldots + X_n^2] + \mathbb{E}[X_1X_2 + X_1X_3 + \cdots + X_{n-1}X_n] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \mathbb{E}[X_iX_j] \\
&= \mathbb{E}[S_n] + (n^2 - n) \cdot \frac{1}{n(n-1)} \\
&= 2
\end{aligned}$$

► *Bruce Lee can break a board with probability 0.8. What is the probability that he breaks at least 2 out of 5 boards, and find a lower bound using Chebyshev's and Hoeffding's inequalities.*

Notate the random variable as B. The probability $P(B \geq 2)$ can be calculated using a simple inversion:

$$1 - \left(\binom{5}{0} 0.8^0 \cdot 0.2^5 + \binom{5}{1} 0.8^1 \cdot 0.2^4\right)$$

The expected value and variance can be calculated:

$$\begin{aligned}
\mathbb{E}[B] &= 0.8 \cdot 5 = 4 \\
\mathrm{Var}(B) &= 5 \cdot 0.8(1 - 0.8) = 0.8
\end{aligned}$$

Now, it can be bounded from below using Chebyshev's Inequality:

$$\begin{aligned}
P(|B - \mathbb{E}[B]| \geq \beta) &\leq \frac{\mathrm{Var}(B)}{\beta^2} \\
P(|B - \mathbb{E}[B]| < \beta) &\geq 1 - \frac{\mathrm{Var}(B)}{\beta^2} \\
P(|B - \mathbb{E}[B]| < 3) &\geq 1 - \frac{0.8}{9} \\
&= 0.911
\end{aligned}$$

To bound it from below using Hoeffding's Inequality:

$$\begin{aligned}
P(B \leq \mathbb{E}[B] - t) &\leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \\
P(B > \mathbb{E}[B] - t) &\geq 1 - \exp\left(-\frac{2t^2}{5}\right) \\
P(B > \mathbb{E}[B] - 3) &\geq 1 - \exp\left(-\frac{18}{5}\right) \\
P(B > 1) = P(B \geq 2) &= 0.973
\end{aligned}$$