Data 140 Lecture Notes

Ryan Lin

Contents

Chapter 1

1.1 Lecture 1

Outcome space: Set of all possible outcomes, denoted by Ω. ω is a single outcome.

Experiment: Roll a die once. (six sided fair die). $\Omega = \{1, 2, 3, 4, 5, 6\}$

Event: Subset of outcome set (A, B, C, etc). Ex: Multiple of 3: $\{3, 6\}$

If the model is "equally likely outcomes", then P(multiple of 3) $= \frac{2}{6} = \frac{\# \text{ event}}{\#\Omega}$

1.1.1 Collisions in Hashing

N codes, where $N \in \mathbb{Z}^+$, where there are n individuals. A collision happens when 2 individuals are assigned the same code.

Model For each individual, pick one code uniformly at random from the N codes, regardless of regardless of all other assignment (independtly).

$$P(\text{no collision}) = \frac{N!/n!}{N^n} = \prod_{i=0}^{n-2} \frac{N-i}{N}$$

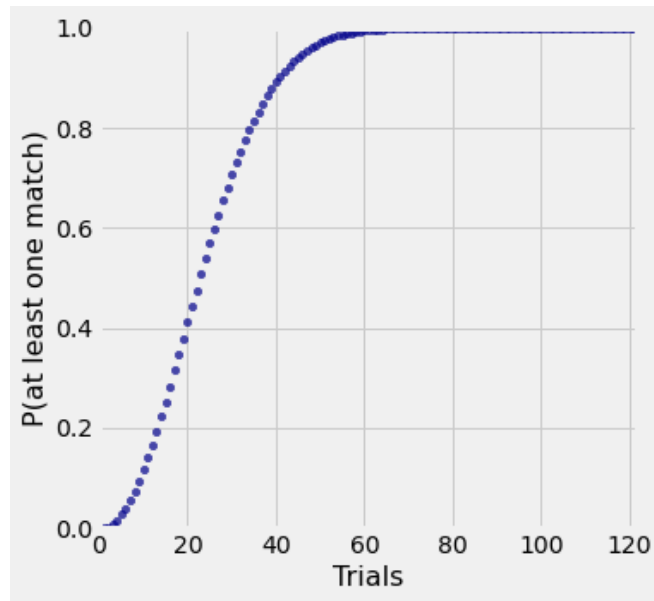$$P(\text{at least one collision}) = 1 - P(\text{no collision}) = 1 - \prod_{i=0}^{n-1} \frac{N-i}{N}$$

A simple way to think about this is human birthdays.

Assumptions:

- Each person is equally likely to be born on each of the 365 days of the year. However, this is not actually true since we have lost leap years and bunching of birth dates, twins, triplets, etc

Under these simplifying assumptions. Compute.

$$\log(P(\text{no collision})) = \sum_{i=0}^{n-1} \log(\frac{N-i}{N})$$

$$= \sum_{i=0}^{n-1} \log(1 - \frac{i}{N})$$

$$\sim \sum_{i} \frac{-i}{N} \approx -\frac{1}{N} \sum i \approx -\frac{1}{N} \frac{(n-1)(n)}{2}$$

For small i, $log(1+x) \sim x$

Chapter 2

2.1   Lecture 02: Rules of Probability

1. Ω outcome space

2. Events are subsets of Ω

3. Probability P is a function on events $P(A)$

Axioms of Probability:

(Ground Rules)

1. $P(A) \geq 0 \forall A$ All probabilities are nonnegative

2. $P(\Omega) = 1$ The whole outcome space should be 1

3. Addition rule (union) $P(\cup_i A_i) = \sum_i P(A_i)$, given that $A_1, A_2, ...$ are mutually exclusive (don't overlap)

Ex: Complements

Taking the inverse of a subset. $P(A^c) + P(A) = P(\Omega)$. these build on the basic axioms

Ex: Unions

If we know the $P(\text{result} = 5) = P(\text{result} \leq 5) - P(\text{result} \leq 4)$

Conditional Probability

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

This works because if you say A occurs, then $\Omega = A$. Conditioning reduces your outcomes space. Then, since A has occurred then the $\omega = A \cap B$.

Ex: Tickets

Say you are drawing 2 tickets out of 3 without replacement. What is the probability that oyu get a blue and a red ticket?

$P(BR) = \frac{1}{6} = \frac{1}{2}$ of $\frac{1}{3} = P(\text{first B}) \cdot P(\text{second R|first B})$

2.1.1   Notation and Language

1. Outcome space: Ω

2. Random Variable: $\Omega \rightarrow \mathbb{R}$

3.

EX: A die rolled 5 times

$\Omega = 6^5$, which is the results of 5 rolls of a die
Statistic S : sum of the 5 numbers.
Function $\Omega \rightarrow \{5, 6, 7, .., 30\}$

If we try to find

$$P(S > 15) = \sum_{k=16}^{30} P(S = k)$$

Which we read as, the chance that the sum exceeds 15 is the sum of the individual chances, where the results are the bounds [16,30].

Let $D_1 =$ first roll, and $D_2 =$ Second roll. Is $D_1 = D_2$?

$$\omega = [3, 1, 3, 4, 2]$$

Clearly, this shows that $D_1 \neq D_2$. They are different random variables. However, the distributions for D_1 and D_2 are the same. each value has the same uniform distribution. Therefore $D_1 \stackrel{d}{=} D_2$ in **distribution**

Random Variables

X, Y are two random variables.

$$P(X = x, Y = y)$$

This means the probability that the random variable X equals a given value x. The comma is an intersection, thus it is $X = x$ and $Y = y$.

If you sum over all possible x and y, you get 1.

$$P(X = 1, Y = 4) = .0625$$

Chapter 3

3.1 Lecture 04: 29 Jan 2026

Finished with basic concepts and methods for calculating probabilities. This lecture will be about applying these concepts in basic models. They are a set of assumptions about data we see in the world. Involve some elements of probabilities.

What makes a good model?

- reflects the setting of the experiment

- do not try to match every detail of that setting. No model that will pick up on everything. Pick up just enough that you can answer interesting questions

To illustrate a model, we have a Population of size N. This population is split into G good and B bad.

Theorem 3.1.1 Hypergeometric Model

Population of size $N = G + B$

Good refers to something you care about, bad is everything else. These are not value statements or implications of morality. It is like Red or not Red.

From this population, get a simple random sample of size n. A SRS shuffle the population and take n. This is the same as sampling n times without replacement. With this we can find something like $P(g$ good elements in a sample$)$

$$P(\text{g good elements in a sample}) = \frac{\binom{G}{g} \cdot \binom{B}{n-g}}{\binom{N}{n}}$$

The denominators is the total possible samples. The numerator is getting how many good elements, where we choose g good elements and have the rest bad. This is why we have the $n - g$. A quick check is that $G + B = N$ and $g + b = n$.

This equation is a hyper-geometric probability function. (N, G, n). the numbers here are the fixed elements in the model. So in this model, we fix the number of elements, the number of good elements, and our sample size.

We would use this model when we draw without replacement, such as drawing from a deck of cards. Then, we also have to define what good means; does it mean that we get aces, we get a two pair, etc?

3.1.1 Sample with replacement:

Example 3.1.1 (Roll a die)

Roll a die 5 time. What is the probability $P(2$ sixes$)$?

Intuitively, the chance of 2 sixes should be small. This is for a sanity check at the end.

We can think about this as 5 possible slots to fill with each trial. One way we can get a good event is

---

66321. we can also assume that probability of one roll is not affected by other roll. We can say that the probability that one happens is

$$\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3$$

Then, we can count how many ways we can get this event; basically, out of 5 places how many times can we choose 2 places? Thus

$$P(2 \text{ sixes} = \binom{5}{2}\left(\frac{5}{6}\right)^2\left(\frac{5}{6}\right)^3$$

Theorem 3.1.2 Binomial Expansion

The new model for this is:
- n independent Success Failure totals (S/F)
- each trial you have probability p of S

$$P(\text{k successes}) = \binom{N}{k}p^k(1-p)^{n-k} \quad k = 0, 1, 2, \ldots, n$$

The key to this formula is the independence of the trials. Any pattern of k successes and n-k failures will have this formula.
This is the Binomial(n,p) distribution. It is used in the case when you have n independent trials, and you are trying to find the probability of k successes.
Checking this formula:

$$P(0 \text{ successes}) = \binom{n}{0}p^0(1-p)^n - 0 = (1-p)^n$$

We want to examine this $\binom{n}{0}$

$$\binom{n}{0} = \frac{n!}{0!(n-0)!}$$

We define 0! = 1 is because of binomials! Define math on how you want to use them lol. The key here is the $\binom{n}{k}$ comes from he binomial expansion and pascal's triangle.

$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k} \tag{3.1}$$

We can take the binomial expansion, the binomial distribution function is $(p - (1-p))^n = 1^n = 1$.

Note:-

If running these calculations in Python, you can either do a sum of probability mass function (pdf), which evaluates how the probability k successes, or cumulative distribution function (pdf), which evaluates the probability you get $\leq k$ successes

3.1.2 How to identify the mode of a discrete distribution

- X binomial(n,p) distribution.
- P(k) = P(X=k)

This interprets to X counting number of successes of n independent success/failure trials, each of which has p probability of success.

$$P(0) = (1 - p)^n$$

$$P(1) = P(0) \cdot \frac{P(1)}{P(0)}$$

$$P(2) = P(1) \cdot \frac{P(2)}{P(1)} = P(1) \cdot R(2).$$

This is called the consecutive odds ratios.

$$R(k) = P(k) \cdot \frac{P(k)}{P(k-1)}, \quad k = 1, 2, \dots, n$$

This is to build the probabilities from the R'S, instead of brute forcing calculations.

In the binomial distribution:

$$R(k) = \frac{\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}}{\frac{n!}{(k-1)!(n-k+1)!} p^{k-1} (1-p)^{n-k+1}}$$

We do this because it is a lot easier to cancel things out.

$$R(k) = \frac{(n-k+1)p}{k(1-p)} = \left(\frac{n+1}{k} - 1 \right) \frac{p}{1-p}$$

As k increases, $R(k)$ decreases in k.

<div style="border:1px solid; padding:8px;">

Note:-

Usually by convention we have $q = 1 - p$

</div>

If $R(k) = 1 \iff R(k) = R(k-1)$, $R(k) < 1 \iff R(k) < R(k-1)$, $R(k) > 1 \iff P(k) > P(k-1)$

this shows us that a decreasing function can only have 1 or 0 times to cross P $= 1$. So the binomial mode is the Largest k so that $P(k) \geq P(k-1)$, because that is when it is rising.

This happens when

$$\left(\frac{n+1}{k} - 1 \right) \frac{p}{q} \geq 1$$

$$\frac{n+1}{k} - 1 \geq \frac{1-p}{p}$$

$$\frac{n+1}{k} \geq \frac{1}{p}$$

$$k \leq (n+1)p$$

Thus, the mode is the integer part of $(n+1)p$

3.1.3   Situation where n large, p_n small

When this happens, the natural thing is that $np_n \to$ number $\mu > 0$

In the binomial distribution

$$P(0) = (1 - p_n)^n \approx e^{-\mu}$$

$$\log(P(0)) = n \log(1 - P_n) \sim -n(p_n) \approx -/mu$$

8

$$P(1) \approx e^{-\mu} \cdot \frac{\mu}{1}$$

Basically, you can show that

$$R(k) \to \frac{\mu}{k} \text{as } n \to \infty$$

This is what is known as the Poisson distribution

Chapter 4

4.1 Lecture 05: 3 February 2026

4.1.1 Extension of the binomial to more than two categories

Example 4.1.1

Let Population be: 10% of A, 30% of B, 60% of C.
Make 10 draws w/ replacement
N_a number of A, N_b, N_c
What is $P(2A, 3B, 5C)$?

$$P(2A, 3B, 5C) = P(N_a = 2, N_b = 3, N_c = 5) = \binom{10}{2} \cdot \binom{8}{3} \cdot \binom{5}{5} \cdot 0.1^2 \cdot .3^3 \cdot .6^5$$

$$\frac{10!}{2!8!} \cdot \frac{8!}{3!5!} 0.1^2 \cdot .3^3 \cdot .6^5 = \frac{10!}{2!3!5!} .1^2 \cdot .3^2 \cdot .6^5$$

The first step is fixing the sequence e.g. $\{AABBBCCCCC\}$ Then, we need to find different ways to get similar sequences. So use 2 slots for As, 3 slots from the 8 remaining, 5 slots from the 5 remaining. The 8! cancels out.

This is called the **multinomial joint distribution**. It is the distribution of multiple variables. It is very close to the binomial distribution, with known number of draws, independence, and same population.

4.1.2 Poisson Distribution

One way this arises is a binomial large n small $p \approx Poisson$ Parameter $\mu = np$. For example, in Binom$(1000, \frac{2}{1000})$. We would expect around 2 successes, which means $\approx Poisson(2)$. The mean of the Poisson distribution is a whole number, and the integer part of μ is a mode.

We should expect a lot of the probability in a Poisson to be near that parameter μ.Which is used as a model for rare events.

When we compare Binom$(1000, \frac{2}{1000})$ and $Poisson(2)$, we can barely tell the difference. This is what it means to be an approximation.

Regardless, the Poisson is still a distribution in its own right, outside of the binomial distribution. It is not always binomial.

Theorem 4.1.1 Poisson Distribution

$$X \text{ has } Poisson(\mu) \text{ distribution}$$

$$P(X = k) = e^{-\mu} \frac{\mu^k}{k!} c \quad k = 0, 1, 2, 3, \ldots$$

We use the Poisson distribution because it models real life and has very nice properties. Therefore we can answer interesting questions simply.

Example 4.1.2 (Sums)

$X \sim Poisson(\mu)$ $Y \sim Poisson(\lambda)$
With X, Y independent.

$$S = X + Y$$

Find the Distribution of S

$$S = 0, 1, 2, 3, \ldots$$

$$P(S = s) = \sum_{k=0}^{s} P(X = k, Y = s - k)) = \sum_{k=0}^{s} e^{\mu} \frac{\mu^k}{k!} \cdot e^{-\lambda} \frac{\lambda^{s-k}}{(s - k)!}$$

> **Note:-**
> General Note: We can't have something on the RHS that is not on the DHS that remains after the calculations, therefore necessitating the sum.

We are now done with probability theory, and now onto algebra!

$$= e^{-(\mu + \lambda)} \frac{1}{s!} \sum_{k=0}^{s} \frac{s!}{k!(s - k)!} \mu^k \lambda^{s-k}$$

$$= e^{-\mu + \lambda} \frac{1}{s!} \sum_{k=0}^{s} \binom{s}{k} \mu^k \lambda^{s-k}$$

$$= e^{-(\mu + \lambda)} \frac{(\mu + \lambda)^s}{s!}$$

.

> **Note:-**
> We multiply by $\frac{s!}{s!}$ to make things easier

This gives us a new Poisson dist with a parameter $\mu + \lambda$. $S \sim Poission(\mu + \lambda)$. It makes intuitive sense because x is around μ and Y is around λ. Therefore S should be around $\mu + \lambda$

For a large n in Poisson(n), e.g. 60, we will get a distribution that is roughly normal. This is because we add something that is very skewed, it becomes roughly normal when it is summed numerous times together.

Theorem 4.1.2 Summation of a Poisson

$$Possion(X_0) + Poission(X_1) + \cdots + Poission(X_n) = Poission(nX)$$

Given that $X_1, X_2, \ldots, X_n$ are independent.

- S number of successes

- F: number of failures

  If the model is Binom(n, p), $F = n - S$. S, F is perfectly dependent.

  If we randomize the number of trials

  If the model that the number of trials is a Poisson(μ) random variable N, then the behavior changes.

  Given $N = n$ S has the Binom(n, p) distribution.

  Given $N = 0$: $S = 0$ with probability 1.

  Distribution of S

  The conditional distribution of S given s is binomial. BUT NOT S.

$$s = 0, 1, 2, \ldots$$

  There is no end to the number of trials because N is a random variable. Therefore, there is no way that S is binomial.

$$P(S = s) = \sum_{n=s}^{\infty} P(N = n, S = s)$$

  Basically, this states that we need n trials starting from s successes, partitioning the event space

$$= \sum_{n=s}^{\infty} P(N = n) P(S = s \mid N = n)$$

$$= \sum_{n=s}^{\infty} e^{-\mu} \frac{\mu^n}{n!} \cdot \frac{n!}{s!(n-s)!} p^s q^{n-s}, \quad q = 1 - p$$

  Probability theory is over....

  We can cancel out the $n!$

$$e^{\mu} \frac{\mu^s p^s}{s!} \sum_{n=s}^{\infty} \frac{\mu^{n-s} q^{n-s}}{(n-s)!}$$

  We split out the $\mu^s = \mu^{n-s} \cdot \mu^s$

$$e^{\mu} \frac{(\mu p)^s}{s!} \sum_{n=s}^{\infty} \frac{(\mu q)^{n-s}}{(n-s)!}$$

$$= e^{-\mu} \frac{(\mu p)^s}{s!} e^{\mu q}$$

$$e^{-\mu(1-s)} \frac{(\mu p)^s}{s!} = e^{\mu p} \frac{(\mu p)^s}{s!}$$

  therefore have $Poisson(\mu p)$

$$P(S = s, F = f) = P(N = s + f, S = s) = e^{-\mu} \frac{\mu^{s+f}}{(s+f)!} \cdot \frac{(s+f)!}{(s!f!)} p^s q^f$$

$$e^{-\mu p} \frac{(\mu p)^s}{s!} \cdot e^{-\mu q} \frac{(\mu q)^f}{f!}$$

Therefore S and F are independent through this property.

Chapter 5

5.1 Lecture 06: 5 February 2026

5.1.1 Expectation

One of the most important parts of probability. When we develop these concepts, we should be thinking about why this concept has a name. Why is expected value defined? Because we refer to it constantly.

A distribution contains complete information about the behavior of our random variables. We can find the $P(X = x)$. Sometimes this is too much information. We only want to understand roughly how big something is occasionally, which often is finding the center of the distribution. In probability, this is the expectation or average.

Definition 5.1.1

Random Variable: X has expected $E[X]$ $X : \Omega \to \mathbb{R}$. Two equivalent definitions.
On the Domain space:

$$E[X] = \sum_{\forall \omega \in \Omega} X(\omega) \cdot P(\omega))$$

On the range space:

$$E[X] = \sum_{\forall x} x \cdot P(X = x)$$

Important: The average is not necessarily a valid value of random variable, but it is in the same units as the random variable. The expectation is a long run average value.

Example 5.1.1 (toss 2 coins)

$$\omega = \{hh, ht, th, tt\}$$

let x be the number of heads. model: fair coin.

$$x(\omega) = \{2, 1, 1, 0\}$$

$$p(\omega) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4}$$

$$x = \{2, 1, 0\}$$

on the domain space:

$$e[x] = 2 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = 1$$

On the range space:

$$E[X] = 2 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 0 \cdot \frac{1}{4} = 1$$

Typically the domain space definition helps us establish facts, and the range space helps us calculate.

5.1.2 Expected Value of Common Distributions

1. Constant: $P(X = c) = 1, E[X] = c$

2. Bern(p). $E[X] = p$.

3. X Uniform $\{1, 2, 3, \ldots, n\}$. $E[X] = \frac{n+1}{2}$

4. $X :$ Poisson(μ) $E[X] = \mu$

5. X: Geom(p) $E[X] = \frac{1}{p}$

Note:-

Bernoulli is very interesting, because we can define I to be an indicator of A .

$$I_A = \begin{cases} 0 & \text{if A doesn't occur} \\ 1 & \text{if A does occur} \end{cases}$$

Then $I_A \sim$ Bern$(P(A))$
Thus $E[I_A] = P(A)$
We have thus shown that any probability is a distribution as well. The expectation of an indicator is the probability of that event.

Example 5.1.2 (Finding expected value of Poisson)

Let $X \sim$ Poisson(μ)

$$E[X] = \sum_{k=0}^{\infty} k \cdot P(X = k)$$

$$= \sum_{k=0}^{\infty} k \cdot e^{-\mu} \frac{\mu^k}{k!}$$

$$= 0 + \sum_{k=1}^{\infty} k \cdot e^{-\mu} \frac{\mu^k}{k!}$$

$$= \sum_{k=1}^{\infty} e^{\mu} \frac{\mu^k}{(k-1)!}$$

$$= e^{-\mu} \mu \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!}$$

$$= e^{-\mu} \cdot \mu \cdot \sum_{k=2}^{\infty} \frac{\mu^k}{k!}$$

$$= e^{-\mu} e^{\mu} \cdot \mu$$

$$= \mu$$

> **Definition 5.1.2: Geometric Distribution**
>
> Often, this is number of times until event X happens.
>
> i.i.d. Bern(p) sequence means independently and identically distributed sequence of 0 and 1s. AKA a sequence of coin tosses that lands heads with probability p.
>
> Let X be the number of trials til the first success. In this class "til" means up to and **until** the success (inclusive).
>
> $$P(X = k) = q^{k-1}p \quad k = 1, 2, 3, \ldots$$
>
> This is called
>
> $$X \sim geometric(p)$$
>
> $$P(X > k) = q^k$$
>
> We can understand this by knowing the definition of Geometric distribution, because any success after k means that everything else were failures.
>
> We can use the tail sum formula to derive the expected value here.
>
> $$E[X] = 1 + q + q^2 + q^3 + \cdots = \frac{1}{1 - (1 - p)} = \frac{1}{p}$$
>
> Alternatively, we can use conditioning! (173 knowledge)
>
> $$E[X] = 1 + p \cdot P(X - 1|x = 0) + q \cdot P(X - 1|x > 1) = 1 + 0 + qE[X] = \frac{1}{p}$$

> **Definition 5.1.3: Tail sum formula for E(Y) with non neg integer valued Y**
>
> Let Y with values $0, 1, 2, \ldots$ $p_i = P(Y = i)$
>
> $$\begin{aligned} E[Y] = p_1 \\ + p_2 + p_2 \\ + p_3 + p_3 + p_3 \\ . \end{aligned} \qquad = P(Y > 0) + P(Y > 1) + P(Y > 2) + \ldots$$
>
> $$E[Y] = \sum_{k=0}^{\infty} P(Y > k)$$

5.1.3 Proprieties of Expectation

Function of X

Case 1.) Linear

Let $Y = aX + b$ Then $E[Y] = aE[X] + b$.

> **Example 5.1.3**
>
> H is the number of heads in 20 tosses (biased coin). T number of tails. Know $E[H] = 6$. What is $E[T]$?.
>
> $$T = 20 - H, E[T] = 20 - E[H]$$

Case 2.) Nonlinear.

Expectation of $Y = g(X)$. $E[Y] = \sum_{\forall x \in X} g(x) \cdot P(X = x)$

Example 5.1.4

$X = \{1, 0, 1\}$ with chances $P(X = k) = \frac{1}{3}$

$$E[X] = 0$$

$Y = X^2$

The range of Y changes with the squared term. Therefore it has $P(Y = 1) = \frac{2}{3}$ and $P(X = 0) = \frac{1}{3}$.

$$E[Y] = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$$

This is the domain space definition.

Example 5.1.5

$$X \sim \text{Poisson}(\mu), E[X] = \mu$$

Find $E[X(X - 1)]$

$$
\begin{aligned}
E[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1) \cdot e^{-mu} \frac{\mu^k}{k!} \\
&= \sum_{k=2}^{\infty} e^{-\mu} \frac{\mu^k}{(k-2)!} \\
&= e^{-\mu} \cdot \mu^2 \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} \\
&= e^{-\mu} \mu^2 e^{\mu} \\
&= \mu^2
\end{aligned}
$$

.