

# Course notes

Probability theory, by topic rather than by lesson: the formulas of each block of the 39 numbered course notes, one worked example per topic and practice with answers. Built from the one lesson filed so far, <https://lessons.ihsan.cc/ece203/01-course-overview/>, and the problem-set solutions filed since; grows as more of the course is filed and built.

## Axioms, sample spaces and counting

Notes 1–4, Ross ch. 1–2. A **sample space**  $S$  is every outcome an experiment could produce; an **event** is a subset  $E \subset S$ , and set algebra does probability’s bookkeeping:  $E \cup F$  is “at least one”,  $EF$  (this course’s shorthand for  $E \cap F$ ) is “both”,  $E^c$  is “not”.  $P$  is any assignment satisfying three axioms:

$$(A1) \ 0 \leq P[E] \leq 1 \quad (A2) \ P[S] = 1 \quad (A3) \ P\left[\bigcup_i E_i\right] = \sum_i P[E_i] \text{ for disjoint } E_i$$

Only A3 has content; the familiar identities are corollaries of it, not extra assumptions:

$$P[\emptyset] = 0, \quad P[E^c] = 1 - P[E], \quad E \subset F \Rightarrow P[E] \leq P[F], \\ P[E \cup F] = P[E] + P[F] - P[EF]$$

where the last one corrects for double-counting the overlap; pushed to  $n$  events the corrections alternate (**inclusion–exclusion**). **De Morgan’s laws**,  $(\bigcap_{i=1}^n E_i)^c = \bigcup_{i=1}^n E_i^c$  and  $(\bigcup_{i=1}^n E_i)^c = \bigcap_{i=1}^n E_i^c$ , are proved by showing each side is a subset of the other:  $x$  misses the intersection exactly when it misses at least one  $E_j$ . Dropping the overlap term gives  $P[A \cup B] \leq P[A] + P[B]$ , since  $P[AB] \geq 0$ , and induction on  $n$  ( $P[\bigcup_{i=1}^n E_i] \leq P[E_n] + P[\bigcup_{i=1}^{n-1} E_i]$ ) gives the **union bound**  $P[\bigcup_{i=1}^n E_i] \leq \sum_{i=1}^n P[E_i]$ . When  $S$  is finite and every outcome is equally likely (an extra modelling assumption, not something the axioms supply),  $P[E] = |E|/|S|$ : probability becomes counting, and the same problem must give the same answer whether outcomes are counted ordered or unordered.

WORKED EXAMPLE · ECE203 TUTORIAL 1 (2021), EXAMPLE T1.4 · CORE · OFFICIAL

An urn has 10 balls numbered 1 to 10. Three are drawn at random. What is the probability that one is even and two are odd?

With order:  $|S_1| = 10 \times 9 \times 8 = 720$ ; the even ball can be first, second or third, each in  $5 \times 5 \times 4 = 100$  ways, so  $300/720 = 5/12$ . Without order:  $|S_2| = \binom{10}{3} = 120$  and  $|E_2| = \binom{5}{2} \binom{5}{1} = 50$ , so  $50/120 = 5/12$  again. Count  $S$  and  $E$  the same way and the factor of 6 cancels.

## Conditioning, Bayes and independence

Notes 5–7, Ross ch. 3. Conditioning renormalises the sample space to what you have learned:

$$P[E | F] = \frac{P[EF]}{P[F]} \quad (P[F] > 0), \quad P[E_1 \cdots E_n] = P[E_1]P[E_2 | E_1] \cdots P[E_n | E_1 \cdots E_{n-1}]$$

Splitting  $S$  into a partition  $F_1, \dots, F_n$  gives the **law of total probability**, run forward from causes to an effect; **Bayes' formula** runs it backward, from the observed effect  $E$  to which cause  $F_j$  produced it:

$$P[E] = \sum_{i=1}^n P[E | F_i] P[F_i], \quad P[F_j | E] = \frac{P[E | F_j] P[F_j]}{\sum_i P[E | F_i] P[F_i]}$$

$P[F_j]$  is the **prior**,  $P[F_j | E]$  the **posterior**; a highly accurate test applied to a rare condition can still leave a positive result more likely a false alarm than a true case, because a small prior dominates a high accuracy. Events are **independent** when  $P[EF] = P[E]P[F]$ , equivalently  $P[E | F] = P[E]$  – a statement about  $P$ , not about the world: mutually exclusive events with positive probability are about as *dependent* as two events get. For three or more events the product rule must hold over every subset, not just the pairs; pairwise independence does not imply it.

Two patterns recur in the problem sets. In **repeated independent rounds** that stop at the first decisive one, condition on that round: if each round independently ends in  $A$  with probability  $a$ , in something else decisive with probability  $b$ , and is replayed otherwise, then  $P[A \text{ ends it}] = \sum_{k \geq 1} (1 - a - b)^{k-1} a = a/(a + b)$ , a geometric series. In **repeated observations of one hidden cause**, the observations are independent *given* the cause, so each likelihood is a product,  $P[E_1 \cdots E_m | F_j] = \prod_i P[E_i | F_j]$ , and goes into Bayes as above. The observations are not independent unconditionally, because each one shifts the odds on the cause.

WORKED EXAMPLE · ECE203 PROBLEM SET 2 (2026), PROBLEM 2 · CORE · OFFICIAL

A bag has balls numbered 1 to 5 and you pick 2, all outcomes equally likely. Given that the product of the two numbers is even, what is the probability that ball 2 was picked?

Order does not change the answer, so count the 10 unordered pairs. The product is even for  $A = \{12, 14, 23, 24, 25, 34, 45\}$ , 7 pairs, and ball 2 is in  $B = \{12, 23, 24, 25\}$ , 4 pairs, all inside  $A$ .

So  $P[B | A] = P[BA]/P[A] = P[B]/P[A] = (4/10)/(7/10) = 4/7$ . Counting the 20 ordered pairs doubles both sets and gives  $8/14 = 4/7$  again.

WORKED EXAMPLE · ECE203 TUTORIAL 2 (2021), EXAMPLE T2.2 · CORE · OFFICIAL

A pouch holds a four-sided and a six-sided die; you pick one at random and roll it. (a) What is the probability of rolling  $i$ ,  $i = 1, \dots, 6$ ? (b) Given a 3, what is the probability you picked the six-sided die?

Let  $A$  be the four-sided die,  $B = A^c$  the six-sided one (each  $1/2$ ),  $E_i$  a roll of  $i$ . (a) Total probability: for  $i \leq 4$ ,  $\frac{1}{4} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} = 5/24$ ; for  $i = 5, 6$ ,  $0 \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} = 2/24$ . (b) Bayes:

$$P[B | E_3] = \frac{\frac{1}{6} \cdot \frac{1}{2}}{\frac{1}{6} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2}} = \frac{2}{5}.$$

## Discrete random variables

Notes 8–14, Ross ch. 4. A random variable is a function  $X : S \rightarrow \mathbb{R}$ . A **discrete** one has pmf  $p_X(a) = P[X = a]$ , and its **expected value** and **LOTUS** are weighted averages against  $p_X$  – in general  $E[g(X)] \neq g(E[X])$ , the course’s most-repeated warning:

$$E[X] = \sum_x x p_X(x), \quad E[g(X)] = \sum_x g(x) p_X(x), \quad \text{Var}[X] = E[X^2] - (E[X])^2$$

and  $E[aX + b] = a E[X] + b$ ,  $\text{Var}[aX + b] = a^2 \text{Var}[X]$ : a shift moves the mean and leaves the spread alone, a scaling stretches the variance by the square. A pmf given up to a constant,  $p_X(n) = C g(n)$ , is fixed by  $\sum_n p_X(n) = 1$ , so  $C = 1/\sum_n g(n)$ ; when the sum is a known series (geometric, or  $\zeta_k = \sum_{n \geq 1} n^{-k}$ ) every moment comes out in the same constants. The  $n$ -th **central moment** is  $\mu_n[X] = E[(X - E[X])^n]$ , so  $\mu_2 = \text{Var}[X]$ ; expand the power and apply linearity to write it through the raw moments  $E[X^k]$ . **Linearity of expectation**,  $E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$ , holds for *any* random variables, independent or not – together with  $E[I_A] = P[A]$  for an indicator, the single most reusable trick in the course. Most problems reduce to recognising one of four families:

TABLE 1 *the four discrete families*

| FAMILY               | COUNTS                  | PMF                                 | $E[X]$    | $\text{Var}[X]$ |
|----------------------|-------------------------|-------------------------------------|-----------|-----------------|
| Bernoulli( $p$ )     | one trial: success?     | $p, 1 - p$                          | $p$       | $p(1 - p)$      |
| Binomial( $n, p$ )   | successes in $n$ trials | $\binom{n}{k} p^k (1 - p)^{n-k}$    | $np$      | $np(1 - p)$     |
| Poisson( $\lambda$ ) | rare events in a window | $\frac{\lambda^k}{k!} e^{-\lambda}$ | $\lambda$ | $\lambda$       |
| Geometric( $p$ )     | trials to first success | $(1 - p)^{k-1} p$                   | $1/p$     | $(1 - p)/p^2$   |

The Poisson is the limit of a binomial as  $n \rightarrow \infty$ ,  $p \rightarrow 0$  with  $np = \lambda$  fixed – many trials, each almost never succeeding, a moderate number expected.

WORKED EXAMPLE · ECE203 TUTORIAL 4 (2021), EXAMPLE T4.2 · CORE · OFFICIAL

$N$ , the number of calls at a call centre in  $t$  seconds, is Poisson with  $\lambda = \alpha t$ ,  $\alpha = 0.2$  calls per second. (i) What is the probability of more than 3 calls in 10 seconds? (ii) Of exactly 12 calls in one minute?

(i)  $\lambda = 0.2 \times 10 = 2$ , so  $P[N > 3] = 1 - e^{-2} \left(1 + 2 + \frac{2^2}{2} + \frac{2^3}{6}\right) = 1 - \frac{19}{3} e^{-2}$ . (ii)  $\lambda = 0.2 \times 60 = 12$ , so  $P[N = 12] = \frac{12^{12}}{12!} e^{-12} \approx 0.11437$ .

## Continuous random variables

Notes 15–21, Ross ch. 5. A random variable is **continuous** when a density  $f_X$  gives area as probability, so  $P[X = a] = 0$  for every single value; only intervals have probabilities. Standardising turns any normal into the standard normal, so one table answers every Gaussian question:

$$P[X \in B] = \int_B f_X(x) dx, \quad \int_{-\infty}^{\infty} f_X dx = 1$$

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1), \quad \Phi(z) = P[Z \leq z], \quad Q(z) = 1 - \Phi(z)$$

The exponential is **memoryless**:  $P[X > s + t \mid X > t] = P[X > s]$  – a component that has already survived  $t$  seconds is as good as new, the only continuous distribution with this property. Given  $Y = g(X)$ , the CDF method ( $F_Y(y) = P[g(X) \leq y]$ , then differentiate) always works; when  $g$  is strictly monotone it collapses to  $f_Y(y) = f_X(g^{-1}(y)) |(g^{-1})'(y)|$ , but a non-monotone  $g$  (e.g.  $Y = X^2$ ) loses a branch under the shortcut and must go through the CDF.

TABLE 2 the three continuous families

| FAMILY                    | PDF                                                        | $E[X]$      | $\text{Var}[X]$ |
|---------------------------|------------------------------------------------------------|-------------|-----------------|
| Uniform( $a, b$ )         | $1/(b - a)$ on $(a, b)$                                    | $(a + b)/2$ | $(b - a)^2/12$  |
| Normal $N(\mu, \sigma^2)$ | $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$ | $\mu$       | $\sigma^2$      |
| Exponential( $\lambda$ )  | $\lambda e^{-\lambda x}$ on $x \geq 0$                     | $1/\lambda$ | $1/\lambda^2$   |

WORKED EXAMPLE · ECE203 MIDTERM, FALL 2023, QUESTION 4 · CORE · OFFICIAL

Let  $X \sim N(10, 4)$ . Find  $a$  such that  $P[aX \leq 1] = 0.15$ , using nearest values in the  $\Phi$  table.

$aX \sim N(a\mu, (a\sigma)^2)$  with  $\mu = 10, \sigma = 2$ . Assuming  $a > 0$ ,  $0.15 = \Phi(b)$  with  $b = (1 - a\mu)/(a\sigma)$ . The table gives  $\Phi(1.04) \approx 0.85$ , so  $b = -1.04$ ,  $1 = a(\mu - 1.04\sigma)$  and  $a = 1/(10 - 1.04 \times 2)$ , positive as assumed.

## Joint distributions

Notes 22–29, Ross ch. 6. The joint CDF  $F_{XY}(a, b) = P[X \leq a, Y \leq b]$  (joint pmf/pdf below it) is the primitive; summing or integrating the other variable away gives the **marginal**:

$$p_X(x) = \sum_j p_{XY}(x, y_j), \quad f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

Marginalisation only runs one way – many joint tables share the same margins.  $X$  and  $Y$  are **independent** exactly when the joint factors into the two marginals ( $F_{XY} = F_X F_Y$ ,  $p_{XY} = p_X p_Y$ , or  $f_{XY} = f_X f_Y$ ); the support must be a rectangle as well as the formula splitting – a joint density that looks like a product on a non-rectangular region (e.g.  $0 < x + y < 1$ ) is a standard trap. For independent  $X, Y$  the density of  $Z = X + Y$  is their **convolution**:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy, \quad p_{X|Y}(x | y) = \frac{p_{XY}(x, y)}{p_Y(y)}, \quad f_{X|Y}(x | y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

A sum of independent normals is normal (means and variances add); a sum of independent Poissons is Poisson (rates add). The **bivariate normal** is parameterised by two means, two variances and  $\rho$ ; its marginals and conditionals are normal, and it is the one place  $\rho = 0$  really does imply independence.

WORKED EXAMPLE · ECE203 TUTORIAL 8 (2021), EXAMPLE T8.2 · CORE · OFFICIAL

$X$  and  $Y$  are independent rolls of a fair four-sided and a fair six-sided die. What is the pmf of  $Z = X + Y$ ?

**Discrete convolution:**  $p_Z(n) = \sum_k p_X(k) p_Y(n-k) = \frac{1}{4} \sum_{k=1}^4 p_Y(n-k)$ , and each surviving term is  $\frac{1}{24}$ , so count how many of  $n-1, \dots, n-4$  lie in  $\{1, \dots, 6\}$ :  $p_Z(n) = \frac{1}{24}, \frac{2}{24}, \frac{3}{24}, \frac{4}{24}, \frac{4}{24}, \frac{3}{24}, \frac{2}{24}, \frac{1}{24}$  for  $n = 2, \dots, 10$ . A box convolved with a longer box is a trapezoid, flat from  $n = 5$  to  $7$ .

## Expectation, generating functions and limit theorems

Notes 30–39, Ross ch. 7–8. Linearity of expectation holds for any pair, however entangled:  $E[X + Y] = E[X] + E[Y]$ . **Covariance** measures the gap that independence closes, and **correlation** normalises it into a fixed range ( $\rho = 0$  does *not* imply independence – an exact functional dependence can still be uncorrelated):

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y], \quad \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X] \text{Var}[Y]}} \in [-1, 1]$$

$$\text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i] + 2 \sum_{i < j} \text{Cov}(X_i, X_j), \quad E[X] = E[E[X | Y]]$$

The last is the **tower property**, the expectation analogue of total probability; the same idea splits a variance as  $\text{Var}[X] = E[\text{Var}(X | Y)] + \text{Var}(E[X | Y])$ . The **MGF**  $M_X(t) = E[e^{tX}]$  recovers moments by differentiating at 0, and turns an independent sum's convolution into a product:  $M_{X+Y}(t) = M_X(t) M_Y(t)$ . **Markov and Chebyshev** need only a mean, or a mean and a variance, and give the **weak law of large numbers**; the **central limit theorem** says how the sample mean is spread around  $\mu$ , independent of what the  $X_i$  were:

$$P[X \geq a] \leq \frac{E[X]}{a} \quad (X \geq 0), \quad P[|X - \mu| \geq b] \leq \frac{\sigma^2}{b^2},$$

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \rightarrow N(0, 1)$$

Chebyshev is a guaranteed bound for every  $n$  and is loose; the CLT is only asymptotic but sharp ( $n \geq 30$  the usual rule of thumb). Inverting the CLT gives a **confidence interval** for  $\mu$  from a sample mean  $\bar{X}$  and sample standard deviation  $S$  ( $t_{\alpha/2} \rightarrow z_{\alpha/2}$  for large  $n$ ), and the **strong law** strengthens the weak law from “rarely far” to “certain to converge”:

$$\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}, \quad P\left[\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu\right] = 1$$

WORKED EXAMPLE · ECE203 FINAL, FALL 2024, QUESTION 7 · CORE · OFFICIAL

One hundred independent  $X_i \sim U(0, 1)$  are generated. You win if  $X_1 + \dots + X_{100} > 55$ . Use the central limit theorem to approximate the probability that you win, expressed with  $\Phi(\cdot)$ .

For  $U(0, 1)$ ,  $E[X_i] = 0.5$  and  $\text{Var}[X_i] = 1/12$ . Centre the sum at  $n\mu = 50$  and scale by  $\sigma\sqrt{n} = \sqrt{100/12} = 10/\sqrt{12}$ , so the event is  $Z_n > 5/(10/\sqrt{12}) = \sqrt{3}$ . By the CLT the answer is  $1 - \Phi(\sqrt{3}) \approx 1 - \Phi(1.7321) \approx 0.0418$ .

## Practice, with answers

- (Tutorial 1, T1.3)** Show that  $P[AB^cC^c] \geq P[A] - P[AB] - P[AC]$ . *Answer:* split  $A$  into four disjoint pieces by  $B$  and  $C$ ;  $ABC \cup AB^cC = AC$  and  $ABC^c \subset AB$ .
- (Problem Set 2, P1, official)** Pick 2 of 5 numbered balls. What is the probability you picked balls 4 and 5, counting (a) unordered and (b) ordered outcomes? *Answer:* (a)  $1/\binom{5}{2} = 1/10$ ; (b)  $2/(5 \times 4) = 1/10$ .
- (Midterm 2024, Q3)** Capacitors are bad with probability 0.1; a good one survives  $t$  seconds with probability  $e^{-at}$ , a bad one  $e^{-bt}$ ,  $b > a > 0$ . What test length  $t_0$  makes 99% of those sold good? *Answer:* Bayes gives  $e^{(b-a)t_0} \geq 0.99/0.09 = 11$ , so  $t_0 \geq \ln 11/(b - a)$ .
- (Own)** Ann and Ben flip together each round, heads with probability  $1/2$  and  $1/3$ . A lone head wins, two heads tie, two tails replay. What is  $P[\text{Ann wins}]$ ? *Answer:* per round Ann alone is  $\frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$ , Ben alone  $\frac{1}{6}$ , a tie  $\frac{1}{6}$ , so  $P = \frac{1}{3}/(\frac{1}{3} + \frac{1}{6} + \frac{1}{6}) = 1/2$ .
- (Own)** A die is fair or loaded ( $P[6] = 1/2$ ), each with probability  $1/2$ . Two independent rolls both show 6. What is the probability the die is loaded? *Answer:*  $(\frac{1}{4} \cdot \frac{1}{2})/(\frac{1}{4} \cdot \frac{1}{2} + \frac{1}{36} \cdot \frac{1}{2}) = 9/10$ .
- (Midterm 2025, Q4)**  $p_X(n) = 1/3^n$  for  $n \geq 1$ ,  $p_X(0) = p_0$ . Find  $p_0$ ,  $E[X]$  and  $E[2^X]$ . *Answer:*  $p_0 = 1/2$ ,  $E[X] = 3/4$ ,  $E[2^X] = 2.5$  (not  $2^{3/4} \approx 1.68$ ).
- (Own)**  $p_X(n) = C/2^n$  for  $n = 1, 2, \dots$ . Find  $C$ ,  $E[X]$ ,  $E[X^2]$  and  $\text{Var}[X]$ . *Answer:*  $C = 1$  (it is Geometric( $\frac{1}{2}$ )),  $E[X] = 2$ ,  $E[X^2] = 6$ ,  $\text{Var}[X] = 6 - 4 = 2$ .
- (Tutorial 6, T6.3)**  $X \sim N(0, 1)$  and  $Z = \sqrt{|X|}$ . Find  $f_Z$ . *Answer:*  $F_Z(a) = 1 - 2\Phi(-a^2)$ , so  $f_Z(a) = \frac{2\sqrt{2}a}{\sqrt{\pi}}e^{-a^4/2}$  for  $a \geq 0$ , and 0 otherwise.
- (Final 2025, Q4)**  $f_{XY} = x + y$  on  $0 < x < a$ ,  $0 < y < b$ . When is it valid? With  $a = b = 1$ , find  $P[X < Y]$  and  $f_{X|Y}$ . *Answer:*  $ab(a+b) = 2$ ;  $P[X < Y] = 1/2$ ;  $f_{X|Y}(x | y) = (x+y)/(\frac{1}{2} + y)$ , which depends on  $y$ , so not independent.
- (Final 2024, Q5)** Start with \$1; draw  $N \sim \text{Poisson}(\lambda)$  and flip a fair coin  $N$  times, a head multiplying by 1.1, a tail by 0.9. Find  $E[Y]$ ,  $\text{Var}[Y]$ . *Answer:*  $E[Y] = 1$ ;  $E[Y^2] = e^{\lambda/100}$ , so  $\text{Var}[Y] = e^{\lambda/100} - 1$ .

Built from lesson 01-course-overview (run 74b04d) and its companion, read 2026-09-27, and the official Problem Set 1 and 2 solutions, read 2026-10-05. The outline bars generative AI from graded work, so Problem Sets 3 and 4 are taught here with own problems, never worked.