STOR 435 Midterm 2 Study Guide

Bayes and independence · Random variables · The famous distributions · Joint distributions and covariance · Normal, limit theorems and inequalities

0. Before you start

WHEN. Thursday, November 12, 2026, in class (5:00–6:15 pm, Gardner 105). Check Canvas in case the date or the stopping point moved.

FORMAT. Closed book and notes. No computer. Bring a calculator that does e^x, factorials and nCr. Expect the style of Midterm 1 and of the posted practice set: multiple choice, mostly "compute a number, pick the interval that contains it", plus a few "pick the correct formula". The syllabus' grading example scores Midterm 2 out of 12, so plan on roughly a dozen questions in 75 minutes.

WEIGHT. 20% of the course grade, or 30% if it turns out to be your better midterm (the syllabus uses whichever of the two grading schemes is higher for you). A strong Midterm 2 can erase a weak Midterm 1.

What is in scope

The syllabus listed *slide decks 5–8, possibly some of 9*. Two things settle it: the instructor has confirmed that **deck 9 is not on the midterm**, and her own *STOR 435 Study Guide: Lectures 5–8* lists exactly what is. That scope is decks 5–8, which runs from conditional probability through the Normal distribution and ends, at the back of deck 8, with the central limit theorem, the law of large numbers and the Markov and Chebyshev inequalities (the syllabus had filed those under deck 9, but the slides and the official guide put them in deck 8). Deck 9 proper is jointly continuous random variables, which is left out. The homework is the best map of what will be asked, because the quizzes and exams follow the homework.

Deck	Classes	Topics	Homework	Guide
5	Sept 10, 15	Law of total probability, Bayes' formula, independence, independent trials, first-step analysis	HW5 (end), HW6	§2–3
6	Sept 24 – Oct 8	Random variables; pmf, cdf, pdf; expectation, variance; moment generating functions; Binomial, Poisson and the Poisson process; Uniform, Exponential	HW7, HW8	§4–10
7	Oct 13, 20	Geometric distribution; joint cdf and pmf, marginals; expectation of sums; indicators (coat matching, coupon collector); multinomial	HW9	§11
8	Oct 22, 27, 29, Nov 3	Independent random variables; sums via convolution and MGFs; covariance, variance of sums, correlation; the Normal distribution; CLT and the normal approximation; law of large numbers; Markov and Chebyshev	HW9, HW10	§12–14

Sources: the filled-in and in-class-annotated slides for decks 5–8, Homeworks 6–10, the posted Midterm 2 practice problems, and the instructor's *Lectures 5–8 Study Guide*. Every result that guide marks with its warning sign is in a blue box here, with the explanation and the worked slide examples the guide itself leaves out.

How to use this guide

- **It assumes you missed class.** Every section starts from the definition and builds up. Blue boxes are the formulas and facts you must know cold. Orange boxes are the mistakes that cost points. Green boxes say how a question is likely to be phrased.
- **Every homework problem is worked here**, inside the section that teaches the idea it needs. The in-class quizzes were variants of the homework, and the exam will be too.
- **Do each "Check yourself" with pen and calculator** before opening the answer. On the exam you need the actual number to four decimals, not just the setup.
- **Section 16** is the posted practice set, fully worked, followed by a second practice exam in the same style. **Section 17** is a one-page distribution table and formula sheet for the night before.
- **The worksheet PDF** is the same text with every answer replaced by a ruled box, and all answers collected in a key at the end. Print it if you learn by writing.

If you have one day

1. Section 1 (30 minutes): make sure the series, integrals and Midterm 1 facts are at your fingertips.
2. Sections 2–3 (45 minutes): total probability, Bayes, independence. Do every Check-yourself.
3. Sections 4–7 (90 minutes): random variables, pmf/cdf/pdf, expectation and variance, MGFs. This is the core vocabulary; everything after it is applications.
4. Sections 8–10 (90 minutes): Binomial, Poisson, Uniform, Exponential. Learn the distribution table in section 17 as you go.
5. Sections 11–14 (90 minutes): geometric, joint distributions, indicators, covariance, Normal, CLT and the inequalities.
6. Section 16 Part A (75 minutes, timed): the posted practice problems. Then read every solution, including the ones you got right.
7. Section 17 before bed and again before the exam.

With two days, add Part B of section 16 on day two and redo every Check-yourself you got wrong.

1. The toolkit: what you need from before

Midterm 2 uses the Midterm 1 material constantly, and from section 5 onward it uses calculus. This section is the minimum. If any line here looks unfamiliar, fix that first; everything later assumes it.

1.1 Probability facts from Midterm 1

THE RULES.

- **Sample space** S is the set of all outcomes; an **event** is a subset. $P(S) = 1$, $0 \le P(E) \le 1$, and for disjoint events probabilities add.
- **Complement:** $P(E^c) = 1 - P(E)$. "At least one" is almost always 1 − P(none).
- **Inclusion–exclusion:** $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.
- **Equally likely outcomes:** P(E) = (number of outcomes in E)/(number in S). Count both at the same level (ordered with ordered, unordered with unordered).
- **Counting:** n! orderings of n things; $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ ways to choose r of n; $n!/(n_1! \cdots n_r!)$ ways to split n things into labeled groups.
- **Conditional probability:** $P(E \mid F) = \frac{P(E \cap F)}{P(F)}$, so $P(E \cap F) = P(F)\,P(E \mid F)$. Given F, the sample space shrinks to F.
- **Multiplication rule for a chain:** $P(E_1 \cap E_2 \cap E_3) = P(E_1)\,P(E_2 \mid E_1)\,P(E_3 \mid E_1 \cap E_2)$.
- **De Morgan:** $(E \cup F)^c = E^c \cap F^c$ and $(E \cap F)^c = E^c \cup F^c$. "Not (at least one)" = "none".

1.2 Series you must recognize on sight

FOUR SERIES.

$$\text{Geometric: } 1 + x + x^2 + \cdots = \frac{1}{1-x} \text{ for } |x| < 1; \text{ starting at } x^k\text{: } \frac{x^k}{1-x}; \text{ finite: } 1 + x + \cdots + x^n = \frac{1 - x^{n+1}}{1-x}.$$

$$\text{Exponential: } \Sigma_{k=0}^{\infty} \frac{x^k}{k!} = e^x \text{ for every } x.$$

$$\text{Binomial theorem: } (a+b)^n = \Sigma_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

$$\text{Limit: } \left(1 + \frac{x}{n}\right)^n \to e^x \text{ as } n \to \infty, \text{ so } \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}.$$

Why these four: the geometric series is how you sum "first success on trial k" probabilities (section 3 and section 11). The exponential series is why the Poisson probabilities add to 1 and where the Poisson MGF comes from (section 9). The binomial theorem is why the Binomial probabilities add to 1 and gives the Binomial MGF (section 8). The limit is how the Binomial turns into the Poisson.

DERIVATIVE TRICK. Differentiating the geometric series gives $\Sigma_{k=1}^{\infty} k\,x^{k-1} = \frac{1}{(1-x)^2}$. That is where the geometric distribution's mean 1/p comes from. You do not have to re-derive it, but recognize it.

1.3 Calculus you must do fast

Derivatives.

- $\frac{d}{dt}\, e^{at} = a\, e^{at}$; $\frac{d}{dt}\, e^{g(t)} = g'(t)\, e^{g(t)}$ (chain rule). Example: $\frac{d}{dt}\, e^{\lambda(e^t - 1)} = \lambda e^t\, e^{\lambda(e^t - 1)}$.
- $\frac{d}{dt}\, (q + pe^t)^n = n(q + pe^t)^{n-1} \cdot pe^t$.
- Product rule: $(uv)' = u'v + uv'$. You need it for second derivatives of MGFs (section 7).
- Fundamental theorem: $\frac{d}{da} \int_{-\infty}^{a} f(x)\, dx = f(a)$. That is why the pdf is the derivative of the cdf (section 5).

Integrals.

- $\int_a^b c\, dx = c(b - a)$. $\int x^n\, dx = \frac{x^{n+1}}{n+1}$ $(n \neq -1)$. $\int_a^\infty x^{-n}\, dx = \frac{a^{1-n}}{n-1}$ for $n > 1$.
- $\int_0^a \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda a}$; $\int_a^\infty \lambda e^{-\lambda x}\, dx = e^{-\lambda a}$.
- Integration by parts, $\int u\, dv = uv - \int v\, du$, gives $\int_0^\infty x\, \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$ and $\int_0^\infty x^2\, \lambda e^{-\lambda x}\, dx = \frac{2}{\lambda^2}$. In general $\int_0^\infty x^n\, e^{-x}\, dx = n!$.
- $\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{(2\pi)}$. This single fact is why the Normal density has the constant it has.
- An even function ($f(-x) = f(x)$) integrates to twice the right half; an odd function integrates to 0 over a symmetric interval.

Worked integral. $\int_0^\infty x\, \lambda e^{-\lambda x}\, dx$ with $u = x$, $dv = \lambda e^{-\lambda x}\, dx$, so $du = dx$, $v = -e^{-\lambda x}$: uv from 0 to ∞ is 0 ($x\, e^{-\lambda x} \to 0$), and $-\int v\, du = \int_0^\infty e^{-\lambda x}\, dx = 1/\lambda$. This is the mean of the Exponential (section 10).

1.4 Calculator habits

- Keep four decimals until the end; round only the final number. Interval answers are usually generous, but (0.25, 0.4] versus (0.4, 0.5] hinges on the fourth decimal sometimes.
- Know where nCr, x!, e^x and ln are on your calculator. $0! = 1$ and $e^0 = 1$.
- Powers like 0.95^{10} are fine on the calculator; 365! is not. Multiply fractions one at a time.
- Have ready: $e^{-1} = 0.3679$, $e^{-2} = 0.1353$, $e^{-3} = 0.0498$, $e^{-4} = 0.0183$, $e^{-5} = 0.0067$, $\ln 2 = 0.6931$, $\sqrt{(2\pi)} = 2.5066$, $\sqrt{(2/\pi)} = 0.7979$.

1.5 Check yourself

1. Sum $\Sigma_{k=3}^{\infty}\, (0.8)^k$.

Starts at $k = 3$: $0.8^3/(1 - 0.8) = 0.512/0.2 = 2.56$.

2. Evaluate $\Sigma_{k=0}^{\infty}\, e^{-2}\, 2^k/k!$ without a calculator.

$e^{-2} \cdot e^2 = 1$. (These are the Poisson(2) probabilities; they add to 1.)

3. Compute $\int_{10}^{\infty} 3000\, x^{-4}\, dx$.

$3000 \cdot 10^{-3}/3 = 3000/3000 = 1$. (So $3000/x^4$ on $x \geq 10$ is a valid density; HW10 problem 3.)

4. Differentiate $M(t) = (0.7 + 0.3e^t)^4$ and evaluate at $t = 0$.

$M'(t) = 4(0.7 + 0.3e^t)^3 \cdot 0.3e^t$; $M'(0) = 4 \cdot 1 \cdot 0.3 = 1.2$. (The mean of a Binomial(4, 0.3).)

5. $P(A) = 0.6$, $P(B \mid A) = 0.5$, $P(B \mid A^c) = 0.2$. Find $P(A \cap B)$ and $P(B)$.

$P(A \cap B) = 0.6 \cdot 0.5 = 0.30$. $P(B) = 0.30 + 0.4 \cdot 0.2 = 0.38$.

6. $\lim_{n\to\infty} (1 - 3/n)^n$ and $\lim_{n\to\infty} (0.97)^n$.

$e^{-3} \approx 0.0498$, and 0 (a fixed base below 1 raised to a growing power).

7. $\int_{-\infty}^{\infty} x\, e^{-x^2/2}\, dx = ?$

0: the integrand is odd and the interval is symmetric. (This is why a standard Normal has mean 0.)

Deck 5 2. Total probability and Bayes' formula

The idea. Many probabilities are easiest to compute by splitting the world into cases ("does she have the disease or not?", "which bus did I take?"), computing the probability inside each case, and averaging with the case probabilities. That is the law of total probability. Bayes' formula is the same computation read backwards: it turns "probability of the evidence given the cause" into "probability of the cause given the evidence", which is what a doctor, a juror or a spam filter actually wants.

2.1 The law of total probability

Two cases. For any events E and F with $0 < P(F) < 1$,

$$P(E) = P(E \cap F) + P(E \cap F^c) = P(E \mid F)\,P(F) + P(E \mid F^c)\,P(F^c).$$

Either F happens or it does not, and the two pieces of E do not overlap.

Many cases (a partition). If $F_1, \ldots, F_n$ are mutually disjoint and together cover S, then

$$P(E) = \Sigma_{i=1}^{n}\, P(E \cap F_i) = \Sigma_{i=1}^{n}\, P(E \mid F_i)\,P(F_i).$$

The two-case formula is $n = 2$ with $F_1 = F$, $F_2 = F^c$.

The recipe the slides use every time: *name the events first, write the formula in symbols, and only then plug in numbers.* Most errors in this topic come from plugging numbers into the wrong slots.

Example: mammograms (slides). In the population described, 0.8% of women have breast cancer. A woman with cancer has a positive mammogram with probability 0.90; a woman without cancer, with probability 0.07. Let C = has cancer, T = positive test. The cases are C and C^c:

$$P(T) = P(T \mid C)\,P(C) + P(T \mid C^c)\,P(C^c) = 0.90(0.008) + 0.07(0.992) = 0.0072 + 0.06944 = 0.07664.$$

About 7.7% of these women test positive, and most of those positives (0.069 of the 0.077) come from women without cancer.

Example: triage (slides). 15% of ER patients are critical, 45% serious, 40% stable. Death rates are 50%, 20% and 2%. With D = dies,

$$P(D) = 0.50(0.15) + 0.20(0.45) + 0.02(0.40) = 0.075 + 0.09 + 0.008 = 0.173, \text{ so } P(\text{survives}) = 0.827.$$

Example: grad school (slides). Eva gets into school A, B or C with probabilities 0.20, 0.40, 0.40, and the chances of a great job J afterwards are 0.95, 0.75, 0.05. Then $P(J) = 0.95(0.20) + 0.75(0.40) + 0.05(0.40) = 0.19 + 0.30 + 0.02 = 0.51$.

Example: the class-selection problem (slides). A class has 24 students; 6 have already been called on. Today a first group of 3 is picked at random, then a second group of 3 is picked at random (from all 24, so it may overlap anything). Find P(N), where N = the second group contains nobody from the original 6 and nobody from the first group.

The difficulty: how many students are "already chosen" when the second group is drawn depends on how much the first group overlapped the original 6. So condition on that. Let A_j = the first group contains exactly j of the original 6, j = 0, 1, 2, 3. These four events partition the sample space.

① $P(A_j) = \dfrac{\binom{6}{j}\binom{18}{3-j}}{\binom{24}{3}}$ (choose j from the 6 and $3 - j$ from the other 18; $\binom{24}{3} = 2024$).

② Given A_j, the number of distinct already-chosen students is $6 + 3 - j = 9 - j$, so the second group must come from the other $15 + j$: $P(N \mid A_j) = \dfrac{\binom{15+j}{3}}{\binom{24}{3}}$.

③ $P(N) = \Sigma_{j=0}{}^{3}\, P(N \mid A_j)\, P(A_j) \approx 0.2649$. So P(at least one overlap) $= 1 - 0.2649 = 0.7351$.

Shortcut check from the slides: choose the second group first. It must avoid the original 6, $\binom{18}{3}$ ways; then the first group can be any 3 of the remaining 21 students, $\binom{21}{3}$ ways; so $P(N) = \binom{18}{3}\binom{21}{3}/\binom{24}{3}^2 = 816 \cdot 1330/2024^2 \approx 0.2649$. Same answer, no conditioning needed. On an exam, use whichever you see first, but know that "condition on the first stage" always works.

How it is asked. "The proportion of policy holders who die next year is in the interval …" or "What proportion survived?" Three groups with three rates: multiply each rate by its group share and add. Watch for the question asking for the complement (survive, not die).

2.2 Bayes' formula: flipping the condition

Bayes' formula.

$$P(F \mid E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E \mid F)\, P(F)}{P(E \mid F)\, P(F) + P(E \mid F^c)\, P(F^c)}.$$

With a partition $F_1, \ldots, F_n$:

$$P(F_j \mid E) = \frac{P(E \mid F_j)\, P(F_j)}{\Sigma_{i=1}{}^{n}\, P(E \mid F_i)\, P(F_i)}.$$

Top: the one case you are asking about. Bottom: the total probability of the evidence E, which is exactly the formula of 2.1.

There is nothing new to memorize: Bayes is the definition of conditional probability, $P(F \mid E) = P(F \cap E)/P(E)$, with the numerator written as $P(E \mid F)\, P(F)$ by the multiplication rule and the denominator expanded by total probability.

Example: mammograms, continued. A woman from this population tests positive. The chance she has cancer is

$$P(C \mid T) = \frac{P(T \mid C)\, P(C)}{P(T)} = \frac{0.90(0.008)}{0.07664} = \frac{0.0072}{0.07664} \approx 0.094.$$

About 9%, even though the test catches 90% of cancers. When Gigerenzer asked 24 doctors this question, their answers ranged from 1% to 90%. The reason the correct answer is small: cancer is rare here, so the 7% false-positive rate applied to the large healthy group produces far more positives than the 90% true-positive rate applied to the small cancer group.

Example: HIV test in a low-risk group (slides, from *Calculated Risks*). P(HIV) = 0.0001; P(positive | HIV) = 0.999; P(negative | no HIV) = 0.9999, so P(positive | no HIV) = 0.0001. Then

$$P(H \mid T) = \frac{0.999(0.0001)}{0.999(0.0001) + 0.0001(0.9999)} = \frac{0.0000999}{0.0000999 + 0.00009999} \approx 0.50.$$

A positive result in this population is a coin flip, which is why a confirmatory test is always run. The slides' story of Betty (told she was HIV positive, two years of medication, then a negative retest and a $600,000 jury award) is the human cost of forgetting this.

Example: grad school, continued. Given that Eva has a great job, the chance she went to A is $\mathrm{P}(\mathrm{A} \mid \mathrm{J}) = \dfrac{0.95(0.20)}{0.51} = \dfrac{0.19}{0.51} \approx 0.373$.

Example: a negative test (practice set, problem 5). 40% of adults are addicted to the Internet. A scan is positive for 70% of addicted adults and for 20% of non-addicted adults. Someone tests *negative*; chance they are addicted? The evidence is now "negative", so use $\mathrm{P}(\text{negative} \mid \text{addicted}) = 0.3$ and $\mathrm{P}(\text{negative} \mid \text{not}) = 0.8$:

$$\mathrm{P}(\text{addicted} \mid \text{negative}) = \frac{0.4(0.3)}{0.4(0.3) + 0.6(0.8)} = \frac{0.12}{0.60} = 0.20.$$

READ WHICH RESULT WAS OBSERVED. If the person tested negative, every conditional probability in the formula must be a probability of a negative result. Using 0.7 and 0.2 here (the positive-result numbers) gives 0.7, which is wrong.

2.3 Natural frequencies: Bayes without the formula

The slides recommend the way doctors are now taught: imagine a large population and count. For the HIV example, take 100,000 low-risk men.

	Have HIV	No HIV	Total
Test positive	≈ 10 (99.9% of 10)	≈ 10 (0.01% of 99,990)	≈ 20
Test negative	≈ 0	≈ 99,980	≈ 99,980
Total	10	99,990	100,000

Among the roughly 20 positives, 10 are real: $\mathrm{P}(\text{HIV} \mid \text{positive}) \approx 10/20 = 50\%$. For the mammogram: 100,000 women, 800 with cancer of whom 720 test positive; 99,200 without cancer of whom 6,944 test positive; $720/(720 + 6{,}944) = 0.094$. Use this to sanity-check any Bayes computation: if the counts and the formula disagree, one of them has a slot filled wrong.

2.4 Two courtroom lessons from the slides

Sally Clark. After two of her sons died of SIDS, a pediatrician testified that the chance of two SIDS deaths in an affluent family was $(1/8500)^2 = 1$ in 72 million. Two things are wrong. Multiplying treats the deaths as independent, ignoring shared genes, environment and medical factors (section 3). Worse, the court was told $\mathrm{P}(\text{two deaths} \mid \text{innocent})$ when the question was $\mathrm{P}(\text{innocent} \mid \text{two deaths})$; a tiny probability of the evidence under one explanation says nothing by itself about the probability of that explanation. She was convicted in 1999 and released in 2003.

O.J. Simpson. The defense argued that fewer than 1 in 2500 battered women are killed by their partner, so battery is a poor predictor of murder. But the woman in question *had been* murdered. Let B = battered, M = murdered, $\mathrm{M_p}$ = murdered by the partner, M^* = murdered by someone else. The relevant quantity is

$$\mathrm{P}(\mathrm{M_p} \mid \mathrm{B} \cap \mathrm{M}) = \frac{\mathrm{P}(\mathrm{M_p} \cap \mathrm{B})}{\mathrm{P}(\mathrm{M_p} \cap \mathrm{B}) + \mathrm{P}(\mathrm{M}^* \cap \mathrm{B})} = \frac{\mathrm{P}(\mathrm{M_p} \mid \mathrm{B})}{\mathrm{P}(\mathrm{M_p} \mid \mathrm{B}) + \mathrm{P}(\mathrm{M}^* \mid \mathrm{B})} = \frac{1/2500}{1/2500 + 5/100{,}000} = \frac{8}{9} \approx 88\%,$$

using the FBI figure that about 5 in 100,000 women are murdered by someone other than a partner. The slides are careful: this is a statement about the calculation, not about the verdict. The exam lesson is the same as in the Clark case: conditioning on *all* the evidence changes the answer completely.

2.5 Check yourself

1. Policy holders are 50% standard (death rate 0.4 next year), 40% preferred (0.2), 10% ultra-preferred (0.1). P(a random policy holder dies next year)? [Practice 7]

0.5(0.4) + 0.4(0.2) + 0.1(0.1) = 0.20 + 0.08 + 0.01 = 0.29, in (0.275, 1.0].

2. John attends class with probability 0.65. If John attends, Joe attends with probability 0.8; if John does not, Joe attends with probability 0.45. P(exactly one of them attends)? [Practice 2]

Exactly one = (John, not Joe) or (not John, Joe): 0.65(0.2) + 0.35(0.45) = 0.13 + 0.1575 = 0.2875, in (0, 0.35].

3. A survey asks "Do you file honestly?" with the instruction: toss a fair coin; heads, answer truthfully; tails, answer NO regardless. 45% answer YES. What proportion files fraudulently? [HW6 P6]

A YES needs heads and an honest filer: P(YES) = 0.5 · P(honest) = 0.45, so P(honest) = 0.9 and 10% file fraudulently.

4. Urn I has 3 red and 2 blue balls; urn II has 1 red and 4 blue. Flip a fair coin to pick an urn, then draw one ball. P(red)? P(urn I | red)?

P(red) = 0.5(0.6) + 0.5(0.2) = 0.4. P(I | red) = 0.5(0.6)/0.4 = 0.75.

5. Machine A makes 50% of a factory's items with 2% defective, B makes 30% with 3% defective, C makes 20% with 5% defective. An item is defective; P(it came from A)?

P(defective) = 0.01 + 0.009 + 0.01 = 0.029. P(A | defective) = 0.01/0.029 ≈ 0.3448.

6. Same HIV test (sensitivity 0.999, false-positive rate 0.0001) in a group where 5% are infected. P(HIV | positive)?

0.999(0.05)/[0.999(0.05) + 0.0001(0.95)] = 0.04995/0.050045 ≈ 0.9981. Same test, different base rate, completely different meaning of a positive result.

7. Two course sections, each equally likely to be emailed. Section 1 is 30% actuarial students, Section 2 is 50%. Actuarial students reply with probability 0.30 (Section 1) or 0.80 (Section 2); nobody else replies. P(reply)? P(Section 1 | reply)?

P(reply) = 0.5(0.3)(0.3) + 0.5(0.5)(0.8) = 0.045 + 0.20 = 0.245. P(Section 1 | reply) = 0.045/0.245 ≈ 0.1837.

8. In the class-selection problem, compute $\mathbf{P(A_0)}$ and $\mathbf{P(N \mid A_0)}$ numerically.

$P(A_0) = \binom{18}{3}/\binom{24}{3} = 816/2024 \approx 0.4032$. $P(N \mid A_0) = \binom{15}{3}/\binom{24}{3} = 455/2024 \approx 0.2248$.

Deck 5 3. Independence and independent trials

The idea. Two events are independent when learning that one happened tells you nothing about the other. Independence is what lets you *multiply* probabilities, and almost every distribution in this course (Binomial, Poisson, Geometric) is built from independent repetitions of a simple experiment. It is a property you check or assume, not a feeling that two things are "unrelated".

3.1 Definition

Independent events. E and F are independent if

$$P(E \cap F) = P(E)\,P(F).$$

When $P(E) > 0$ and $P(F) > 0$ this is the same as $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$: conditioning on one does not change the other.

Why the two versions agree: if $P(E \cap F) = P(E)P(F)$ then $P(E \mid F) = P(E \cap F)/P(F) = P(E)P(F)/P(F) = P(E)$. Conversely, multiply $P(E \mid F) = P(E)$ by $P(F)$. The product form is the official definition because it also works when a probability is 0.

- **Two coin tosses.** E = first is H, F = second is T. S = {HH, HT, TH, TT}; $E \cap F$ = {HT}; $P(E \cap F) = 1/4 = (1/2)(1/2)$. Independent.
- **One card.** E = club, F = a 6. $P(E \cap F) = 1/52 = (1/4)(1/13)$. Independent: the suit tells you nothing about the rank.
- **Two dice, F = first die is 4 (P = 1/6).** E_1 = sum is 6: $P(E_1) = 5/36$ and $E_1 \cap F = \{(4, 2)\}$ has probability 1/36, but $(5/36)(1/6) = 5/216 \neq 1/36$. *Not* independent. E_2 = sum is 7: $P(E_2) = 6/36 = 1/6$, $E_2 \cap F = \{(4, 3)\}$, and $(1/6)(1/6) = 1/36$. Independent. Changing the target from 6 to 7 changes the answer; you must check.

How it is asked. "E = sum is 7, F = red die shows 4. Then E and F are: disjoint / independent / neither." Compute $P(E \cap F)$ and compare with $P(E)P(F)$. Here $1/36 = (1/6)(1/6)$: independent (practice set, problem 1).

3.2 Complements, and the disjoint trap

Complements inherit independence. If E and F are independent, so are E and F^c, E^c and F, and E^c and F^c. Proof: $P(E \cap F^c) = P(E) - P(E \cap F) = P(E) - P(E)P(F) = P(E)(1 - P(F)) = P(E)P(F^c)$.

Disjoint is the opposite of independent. Disjoint means $E \cap F = \varnothing$, so $P(E \cap F) = 0$. If both events have positive probability, disjoint events can never be independent: if one happens, the other is now impossible, so learning one changes the other. HW6 problems 1 and 2: "first toss H" and "first toss T" are disjoint; "first toss H" and "second toss T" are independent. Pick one; they are not both true.

3.3 Three or more events

Independence of E, F, G. All four equations must hold:

$$P(E \cap F) = P(E)P(F),\ P(E \cap G) = P(E)P(G),\ P(F \cap G) = P(F)P(G),\ P(E \cap F \cap G) = P(E)P(F)P(G).$$

For n events: the probability of every sub-collection's intersection is the product of the probabilities. Then any events built from *disjoint groups* of them are independent: E and F ∪ G; E^c, F^c and G^c; $E_1 \cap E_2$ and ${E_3}^c \cap E_4$; and so on.

Why E is independent of F ∪ G (slides): $P(E \cap (F \cup G)) = P(E \cap F) + P(E \cap G) - P(E \cap F \cap G) = P(E)[P(F) + P(G) - P(F)P(G)] = P(E)P(F \cup G)$.

HW6 problem 3. E_1, E_2, E_3 independent. $P(E_1 \mid E_2 \cap E_3) = P(E_1)$, because E_1 is independent of $E_2 \cap E_3$ (built from the other two).

HW6 problem 4. A, B, C independent with P(A) = 2/3, P(B) = 3/4, P(C) = 1/5.

- $P(A \cap B^c \cap C^c) = (2/3)(1/4)(4/5) = 8/60 = 2/15 \approx 0.1333$.
- P(exactly one occurs) = $P(A\ B^c\ C^c) + P(A^c\ B\ C^c) + P(A^c\ B^c\ C) = 8/60 + (1/3)(3/4)(4/5) + (1/3)(1/4)(1/5) = 8/60 + 12/60 + 1/60 = 21/60 = 0.35$.

HW6 problem 5. E, F, G independent with probabilities 0.2, 0.3, 0.1. $P(E \cup F^c \cup G) = 1 - P(E^c \cap F \cap G^c) = 1 - (0.8)(0.3)(0.9) = 1 - 0.216 = 0.784$. Unions of independent events: go through the complement (De Morgan), then multiply.

Practice set, problem 4. E, F, G, H, I independent with probabilities 0.2, 0.3, 0.1, 0.4, 0.5. $P((E \cup F^c \cup G) \cap H \mid I)$: the event is built from E, F, G, H only, so it is independent of I and the conditioning does nothing. Then $(E \cup F^c \cup G)$ and H are independent (disjoint groups), so the answer is $0.784 \times 0.4 = 0.3136$, in (0.25, 0.5].

3.4 Independent trials

Trials. An experiment made of repeated identical sub-experiments (same outcomes, same probabilities) whose results do not influence each other: coin tosses, dice rolls, items on an assembly line. Any event determined by trial i is independent of any event determined by other trials, so probabilities of specific sequences multiply.

HW6 problem 7. A coin with P(H) = 1/3 is flipped 10 times; E_i = toss i is H. The event "first 6 heads, next 4 tails" is $F = E_1 \cap \cdots \cap E_6 \cap {E_7}^c \cap \cdots \cap {E_{10}}^c$, and $P(F) = (1/3)^6(2/3)^4 \approx 0.00027$.

HW6 problem 8(1). Eva botches each order with probability 1/2, independently. P(first 10 correct, 11th wrong) = $(1/2)^{10}(1/2) = (1/2)^{11}$. (This is the Geometric distribution of section 11 in disguise.)

HW6 problem 8(2): the race. Tim errs with probability 1/6, Eva with 1/2, per customer, independently. Eva takes the trip if she errs strictly before Tim does (a tie means the game stops and Eva stays). For Eva to lose on customer n, both must be perfect on customers 1 to n − 1, probability $[(1/2)(5/6)]^{n-1} = (5/12)^{n-1}$, then Eva errs and Tim does not, probability (1/2)(5/6) = 5/12:

$$P(\text{Eva takes the trip}) = \Sigma_{n=1}^{\infty}\ (5/12)^{n-1}\ (5/12) = \frac{5/12}{1 - 5/12} = \frac{5}{7} \approx 0.7143.$$

By the same method P(Tim first) = (1/12)/(7/12) = 1/7 and P(tie) = (1/12)/(7/12) = 1/7, and 5/7 + 1/7 + 1/7 = 1. Any "who fails first" question is a geometric series: (probability nothing decisive happens)$^{n-1}$ × (probability of the decisive outcome).

HW6 problem 9. Roll two dice repeatedly; E_n = no 3 or 5 on the first n − 1 rolls and a 3 on roll n. P(sum 3) = 2/36 = 1/18, P(sum 5) = 4/36, P(neither) = 1 − 6/36 = 5/6. So $P(E_n) = (5/6)^{n-1}(1/18)$. E_4 and E_5 are disjoint (E_5 requires no 3 on roll 4, E_4 requires a 3 there), so $P(E_4 \cup E_5) = (5/6)^3(1/18) + (5/6)^4(1/18) \approx 0.0589$: option (E). Options with a subtracted product are wrong because there is no overlap to remove, and options with factors 3 and 4 in front count the trials incorrectly.

Example: the flat tire (slides). Four friends independently name which of four tires went flat. Donatello and Michelangelo pick uniformly (1/4 each). Leonardo picks each left tire with twice the probability of each right tire: 2q + 2q + q + q = 1 gives q = 1/6, so left tires 1/3 each, right tires 1/6 each. Raphael is the mirror image. For any fixed tire the four answers agree with probability (1/4)(1/4)(1/3)(1/6) = 1/288, because one biased friend contributes 1/3 and the other 1/6 whichever tire it is. Four tires: P(all match) = 4/288 = 1/72. The structure the slides stress: *first a sum over disjoint cases, then independence inside each case.*

3.5 First-step analysis: "which happens first?"

Example: 5 before 7 (slides). Roll two dice repeatedly. P(a sum of 5 appears before a sum of 7)? Let p be this probability and condition on the first roll: sum 5 (probability 4/36), sum 7 (6/36), or neither (26/36). If neither, the game restarts from scratch, so the conditional probability is p again:

$$p = 1 \cdot \frac{4}{36} + 0 \cdot \frac{6}{36} + p \cdot \frac{26}{36}, \text{ so } p\left(1 - \frac{26}{36}\right) = \frac{4}{36}, \; p = \frac{4}{10} = \frac{2}{5}.$$

SHORTCUT. Only decisive rolls matter: $P(\text{A before B}) = \dfrac{P(\text{A on one roll})}{P(\text{A on one roll}) + P(\text{B on one roll})}$. Here (4/36)/(10/36) = 2/5.

Alternating turns (practice set, problem 3). A rolls a 4-sided and a 6-sided die aiming for sum 3 (outcomes (1,2), (2,1): $p_A = 2/24 = 1/12$); then B rolls aiming for sum 4 ((1,3), (2,2), (3,1): $p_B = 3/24 = 1/8$); they alternate until someone succeeds. Condition on the first round: A wins immediately with probability p_A; otherwise, if B also fails, the game restarts:

$$p = p_A + (1 - p_A)(1 - p_B)\, p, \text{ so } p = \frac{p_A}{1 - (1 - p_A)(1 - p_B)} = \frac{1/12}{1 - (11/12)(7/8)} = \frac{1/12}{19/96} = \frac{8}{19}$$
$$\approx 0.4211,$$

in (0.25, 0.45]. Going first is worth something but A's target is harder, so A is still the underdog.

TWO WAYS TO SEE THE SAME NUMBER. The first-step equation and the geometric series always agree: $p = p_A[1 + r + r^2 + \cdots]$ with $r = (1 - p_A)(1 - p_B)$ gives $p_A/(1 - r)$. Use the equation when the game restarts after an indecisive step; use the series when you want to see the trials.

3.6 Check yourself

1. E and F independent, P(E) = 0.4, P(F) = 0.5. Find P(E ∪ F), P(neither), and P(exactly one).

$P(E \cup F) = 0.4 + 0.5 - 0.2 = 0.7$. $P(E^c \cap F^c) = 0.6 \cdot 0.5 = 0.3$ (= 1 − 0.7). Exactly one: $0.4 \cdot 0.5 + 0.6 \cdot 0.5 = 0.5$.

2. Toss a fair coin twice. E = first toss is H, F = first toss is T. Independent or disjoint? What if F = second toss is T? [HW6 P1–2]

First pair: disjoint ($P(E \cap F) = 0 \neq 1/4$). Second pair: independent ($1/4 = 1/2 \cdot 1/2$).

3. A, B, C independent with P(A) = 2/3, P(B) = 3/4, P(C) = 1/5. $P(A \cap B^c \cap C^c)$ and P(exactly one occurs)? [HW6 P4]

$2/15 \approx 0.1333$ and $21/60 = 0.35$ (worked in 3.3).

4. Three components work independently, each with probability 0.9. P(at least one works)? P(all three work)? P(exactly two work)?

$1 - 0.1^3 = 0.999$; $0.9^3 = 0.729$; $3 \cdot 0.9^2 \cdot 0.1 = 0.243$.

5. E, F, G independent with probabilities 0.2, 0.3, 0.1. $P(E \cup F^c \cup G)$? [HW6 P5]

$1 - P(E^c \cap F \cap G^c) = 1 - 0.8 \cdot 0.3 \cdot 0.9 = 0.784$.

6. Roll two dice repeatedly. P(a sum of 4 appears before a sum of 8)?

$P(4) = 3/36$, $P(8) = 5/36$: $3/(3 + 5) = 0.375$.

7. In the Eva–Tim race, P(the game ends on customer 3)?

Both perfect twice, then at least one errs: $(5/12)^2 (1 - 5/12) = (25/144)(7/12) \approx 0.1013$.

8. A coin with P(H) = 1/3 is tossed 5 times. P(exactly 2 heads)?

Each specific sequence with 2 heads has probability $(1/3)^2(2/3)^3$; there are $\binom{5}{2} = 10$ of them: $10 \cdot (1/9)(8/27) = 80/243 \approx 0.3292$. This is the Binomial pmf of section 8.

9. Players A and B alternate, A first, each succeeding on a turn with probability 1/2. P(A wins)?

$p = (1/2)/(1 - (1/2)(1/2)) = (1/2)/(3/4) = 2/3$.

10. E_n = no 3 or 5 on the first n − 1 rolls of two dice and a 3 on roll n. Which formula gives $P(E_4 \cup E_5)$? [HW6 P9]

$(5/6)^3(1/18) + (5/6)^4(1/18) \approx 0.0589$; the events are disjoint so nothing is subtracted.

Deck 6 4. Random variables: pmf and cdf

The idea. A random variable is a number produced by a random experiment: the number of heads in three tosses, the largest ball drawn, the price of a stock next Monday at 9 a.m. The sample space might be hopelessly complicated (the state of the universe next Monday), but we only care about one measurement of it. From now on the course describes randomness through random variables and their distributions rather than through events and sample spaces, although every statement still is one.

4.1 Definition and the two types

RANDOM VARIABLE. A real-valued function defined on the sample space: it assigns a number X(s) to every outcome s. Notation X, Y, Z. The event "X = 2" is the set of outcomes s with X(s) = 2, so P(X = 2) makes sense.

- **Example 1a.** Toss 3 fair coins; Y = number of heads. Y takes values {0, 1, 2, 3}.
- **Example 1b.** Draw 3 balls without replacement from balls numbered 1 to 20; X = the largest number drawn. X takes values {3, 4, …, 20} (the largest of three distinct numbers is at least 3).
- **Example 1c.** Pick a point "at random" from [0, 1]; X = the point. X takes uncountably many values, and we will need a new way to describe it.

DISCRETE VS CONTINUOUS. A **discrete** random variable takes at most countably many values (a finite list, or a sequence like 0, 1, 2, …). You handle it with sums and counting. A **continuous** random variable takes values in an interval; you handle it with integrals (section 5).

4.2 The probability mass function (discrete)

PMF. For a discrete X taking values x_1, x_2, …,

$$p(a) = P(X = a), \text{ with } p(x_i) \geq 0,\ p(x) = 0 \text{ for every other } x, \text{ and } \Sigma_i\, p(x_i) = 1.$$

The pmf characterizes X: to describe a discrete random variable, list its values and their probabilities. (From the margin of the slides: only countably many points can have $p(x) > 0$.)

Example 1a worked. The 8 equally likely outcomes are HHH, HHT, HTH, THH, HTT, THT, TTH, TTT with Y = 3, 2, 2, 2, 1, 1, 1, 0. So p(0) = 1/8, p(1) = 3/8, p(2) = 3/8, p(3) = 1/8. Check: they add to 1.

Example 1b worked. All $\binom{20}{3} = 1140$ sets of three balls are equally likely. For the largest to equal k, ball k must be in the set and the other two must come from {1, …, k − 1}: $\binom{k-1}{2}$ sets. So

$$P(X = k) = \frac{\binom{k-1}{2}}{\binom{20}{3}},\ k = 3, 4, \ldots, 20.$$

For instance $P(X = 20) = \binom{19}{2}/1140 = 171/1140 = 0.15$, and P(X = 3) = 1/1140.

HW7 problem 2: the second largest. Same urn, X = the second largest of the three numbers, values 2 to 19. For X = i you need one ball below i (i − 1 choices) and one above (20 − i choices):

$$p(i) = \frac{(i-1)(20-i)}{\binom{20}{3}},\ i = 2, \ldots, 19\text{: option (D).}$$

The wrong options have (i + 1) or (20 − i + 1), which would allow a ball to equal i twice. Sanity-check a formula at an edge: $p(19) = 18 \cdot 1/1140$ (one ball must be ball 20, the other anything below 19); $p(2) = 1 \cdot 18/1140$.

Biased coin (slides). Three tosses, $P(H) = 1/4$ on each, independent. Y = number of heads. Each sequence with k heads has probability $(1/4)^k(3/4)^{3-k}$ and there are $\binom{3}{k}$ such sequences:

k	0	1	2	3
$p(k) = \binom{3}{k}(1/4)^k(3/4)^{3-k}$	27/64	27/64	9/64	1/64

This is the Binomial(3, 1/4) distribution of section 8. Notice that "0 heads" and "1 head" are equally likely here, and that the values 0, 1, 2, 3 are certainly not equally likely.

HW7 problem 3 (a table). X takes values −1.3, 1.2, 2.7, 3.3, 9.8 with probabilities 0.1, 0.25, 0.35, 0.1, 0.2. A pmf given as a table is the most common exam format; the questions then ask for probabilities of ranges, the cdf at a point, or expectations (section 6).

4.3 The cumulative distribution function

CDF. For any random variable (discrete or continuous),

$$F(x) = P(X \le x), \quad -\infty < x < \infty.$$

F is non-decreasing, $F(x) \to 0$ as $x \to -\infty$, $F(x) \to 1$ as $x \to \infty$, and F is right-continuous. For a discrete X, F is a staircase: flat between the possible values, jumping at each value x_i by exactly $p(x_i)$. For $a < b$, $P(a < X \le b) = F(b) - F(a)$.

For the biased coin: $F(x) = 0$ for $x < 0$, 27/64 on [0, 1), 54/64 on [1, 2), 63/64 on [2, 3), and 1 for $x \ge 3$. The value *at* a jump is the upper level ($F(1) = 54/64$ includes $P(X = 1)$).

From the cdf back to the pmf.

1. The locations of the jumps are the values X takes with positive probability.
2. The size of each jump is the probability of that value.

HW7 problem 1. $F(x) = 0$ for $x < -2.5$; 2/12 on [−2.5, 0); 5/12 on [0, 1); 8/12 on [1, 6.5); 1 for $x \ge 6.5$. Jumps: at −2.5 of size 2/12, at 0 of size 3/12, at 1 of size 3/12, at 6.5 of size 4/12. So the pmf is

x	−2.5	0	1	6.5
p(x)	2/12	3/12	3/12	4/12

HW7 problem 3(a). For the table above, $F(-1.5) = P(X \le -1.5) = 0$, because the smallest value, −1.3, is larger than −1.5 (nothing sits at or below −1.5). And $F(3.5) = 0.1 + 0.25 + 0.35 + 0.1 = 0.8$: everything at or below 3.5, which excludes only the value 9.8. The cdf at a point that is not a possible value is just the running total of the pmf up to that point.

Endpoints in a discrete range. $P(1 \le X \le 2)$ is *not* $F(2) - F(1)$. It is $P(X = 1) + P(X = 2) = F(2) - F(0)$ for integer values; in general $F(b) - F(\text{just below } a)$. For the biased coin, $P(1 \le Y \le 2) = 63/64 - 27/64 = 36/64 = 9/16$. The slides' solution writes $F(2) - F(1)$ by accident but computes $63/64 - 27/64$; the numbers are what count. Always decide whether the left endpoint is included.

Practice set, problem 8 (reading a cdf graph). The graph of F is 0 left of -2, 0.1 on $[-2, -0.75)$, 0.3 on $[-0.75, 0)$, 0.6 on $[0, 1)$, 0.8 on $[1, 2)$, and 1 from 2 on. Jumps: 0.1 at -2, 0.2 at -0.75, 0.3 at 0, 0.2 at 1, 0.2 at 2. That is the pmf; the expected value (section 6) is then $(-2)(0.1) + (-0.75)(0.2) + 0(0.3) + 1(0.2) + 2(0.2) = -0.2 - 0.15 + 0.2 + 0.4 = 0.25$, in $(0.2, 0.3]$.

4.4 Indicator random variables

INDICATOR OF AN EVENT. $I_A = 1$ if A occurs and 0 if A^c occurs. Its pmf is $P(I_A = 1) = P(A)$, $P(I_A = 0) = 1 - P(A)$. It is the simplest random variable, and (section 11) adding indicators is the fastest way to count how many of several events happen.

4.5 Check yourself

1. Roll two fair dice; X = the larger of the two numbers. Find the pmf of X.

$P(X \le k) = (k/6)^2$ (both dice at most k), so $P(X = k) = (k^2 - (k-1)^2)/36 = (2k-1)/36$: 1/36, 3/36, 5/36, 7/36, 9/36, 11/36 for k = 1, …, 6.

2. A cdf is 0 for $x < -2.5$, 2/12 on $[-2.5, 0)$, 5/12 on $[0, 1)$, 8/12 on $[1, 6.5)$, 1 for $x \ge 6.5$. Give the pmf. [HW7 P1]

$p(-2.5) = 2/12$, $p(0) = 3/12$, $p(1) = 3/12$, $p(6.5) = 4/12$.

3. Draw 3 balls without replacement from 20 numbered balls; X = the second largest. Which formula is p(i)? [HW7 P2]

$(i-1)(20-i)/\binom{20}{3}$ for i = 2, …, 19: one ball below i, one above.

4. For the biased-coin Y (3 tosses, P(H) = 1/4), find F(1.5), P(Y ≥ 2) and P(0 < Y ≤ 2).

$F(1.5) = F(1) = 54/64$. $P(Y \ge 2) = 10/64$. $P(0 < Y \le 2) = F(2) - F(0) = 63/64 - 27/64 = 36/64$.

5. A fair coin is tossed until the first head; X = number of tosses. Find P(X = k), P(X ≤ 3) and P(X > 3).

$P(X = k) = (1/2)^{k-1}(1/2) = (1/2)^k$. $P(X \le 3) = 1/2 + 1/4 + 1/8 = 7/8$; $P(X > 3) = 1/8$ (three tails in a row).

6. Two cards are drawn without replacement; X = number of aces. Find the pmf.

$P(X = 0) = \binom{48}{2}/\binom{52}{2} = 1128/1326 \approx 0.8507$; $P(X = 1) = 4 \cdot 48/1326 \approx 0.1448$; $P(X = 2) = 6/1326 \approx 0.0045$.

7. X takes values −1.3, 1.2, 2.7, 3.3, 9.8 with probabilities 0.1, 0.25, 0.35, 0.1, 0.2. Find F(1.2), F(3.5) and P(1 < X ≤ 3.3). [HW7 P3]

$F(1.2) = 0.1 + 0.25 = 0.35$ (the jump at 1.2 is included). $F(3.5) = 0.8$. $P(1 < X \le 3.3) = F(3.3) - F(1) = 0.8 - 0.1 = 0.7$.

8. Brad's interviews: X = 0, 1, 2, 3 with probabilities 0.1, 0.2, 0.6, 0.1. P(X ≥ 2)? [HW7 P5]

$0.6 + 0.1 = 0.7$.

Deck 6 5. Continuous random variables: pdf and cdf

The idea. A point chosen at random from [0, 1] has uncountably many possible values, so we cannot give each value a positive probability (they would add up to infinity). Instead we describe how probability is *spread* along the line with a density function, and probabilities of intervals are areas under it. Every discrete formula with a sum has a continuous twin with an integral.

5.1 The probability density function

PDF. X is continuous if there is a non-negative function f such that for every set B,

$$P(X \in B) = \int_B f(x)\,dx.$$

In particular $\int_{-\infty}^{\infty} f(x)\,dx = 1$, $P(a \le X \le b) = \int_a^b f(x)\,dx$, and $P(X = a) = \int_a^a f = 0$.

What f(a) means. For a small ε, $P(a - \varepsilon/2 \le X \le a + \varepsilon/2) \approx f(a) \cdot \varepsilon$. So f(a) measures how likely X is to be *near* a; it is not a probability and can be larger than 1 (a density of 2 on [0, 1/2] is fine).

Consequences of $P(X = a) = 0$. For a continuous X, $P(X < a) = P(X \le a)$, $P(a < X < b) = P(a \le X \le b)$, and $P(|X| = 1/2) = 0$ (HW10 problem 6). Endpoints never matter for continuous variables; they always matter for discrete ones.

Uniform on [0, 1] (slides). We want every point in [0, 1] equally likely, so the density is a constant C on [0, 1] and 0 elsewhere. Since $1 = \int_0^1 C\,dx = C$, the density is f(x) = 1 on [0, 1]. Then

- $P(X > 0.25) = \int_{0.25}^1 1\,dx = 0.75$.
- $P(0.5 \le X \le 1.7) = \int_{0.5}^1 1\,dx = 0.5$ (the density is 0 above 1).
- $P(X = 0.787) = 0$.

For a uniform variable, probability = length of the piece of the interval you are asking about, divided by the total length.

5.2 The cdf of a continuous variable

CDF AND PDF.

$$F(a) = P(X \le a) = \int_{-\infty}^a f(x)\,dx, \text{ and } F'(a) = f(a) \text{ wherever F is differentiable.}$$

F is a continuous function (no jumps, since no single value has positive probability). To go from cdf to pdf, differentiate; from pdf to cdf, integrate from $-\infty$ (in practice from the left end of the support).

Uniform on [0, 1]. F(x) = 0 for $x < 0$, $F(x) = \int_0^x 1\,dt = x$ for $0 \le x \le 1$, and F(x) = 1 for $x > 1$. Differentiating the middle piece gives back f = 1. A continuous, non-step cdf is the signature of a continuous random variable.

Tail probabilities. $P(X > a) = 1 - F(a) = \int_a^\infty f$. For densities that live on $[a, \infty)$ the tail integral is often easier than the cdf.

5.3 Finding the constant, then the probabilities

A density is often given as "c · (something)". The first step is always the same: set the total integral equal to 1 and solve for c. Three homework examples.

HW10 problem 3 (a power-law density). $f(x) = c/x^4$ for $x \geq 10$, else 0. These densities model wealth, city sizes and social networks.

1. $1 = \int_{10}^{\infty} c\, x^{-4}\, dx = c \cdot \frac{10^{-3}}{3} = \frac{c}{3000}$, so $c = 3000$.
2. $P(X > 20) = \int_{20}^{\infty} 3000\, x^{-4}\, dx = 3000 \cdot \frac{20^{-3}}{3} = \frac{1000}{8000} = 0.125$.

(The mean and variance of this X are in section 6.)

HW10 problem 6 (a symmetric density). $f(x) = c\, x^2$ on $[-4, 4]$.

1. $1 = c \int_{-4}^{4} x^2\, dx = c \cdot \frac{2 \cdot 64}{3} = \frac{128c}{3}$, so $c = 3/128$.
2. $P(|X| = 1/2) = 0$ (a single value for a continuous variable, or rather two values, $\pm 1/2$, each of probability 0).
3. $P(|X| > 2) = P(X > 2) + P(X < -2) = 2 \cdot \frac{3}{128} \int_2^4 x^2\, dx = 2 \cdot \frac{3}{128} \cdot \frac{64 - 8}{3} = \frac{112}{128} = 0.875$.

HW10 problem 5 (the Laplace density). $f(x) = \frac{1}{2} e^{-|x|}$ on the whole line: two Exponential(1) halves glued back to back, each carrying probability 1/2.

1. $P(|X| \leq 2) = 2 \int_0^2 \frac{1}{2} e^{-x}\, dx = 1 - e^{-2} \approx 0.8647$.
2. $\{|X| \leq 2 \text{ or } X \leq -1\} = \{X \leq 2\}$ (the second set swallows the left part), so the probability is $P(X \leq 2) = 1 - P(X > 2) = 1 - \frac{1}{2} e^{-2} \approx 0.9323$.
3. $P(X = 1 \text{ or } X = 2 \text{ or } \ldots) = 0$: a countable union of probability-zero events.

> SYMMETRY. If f is even, $P(X > a) = P(X < -a)$ and $P(|X| > a) = 2\, P(X > a)$. It halves the work.

5.4 Conditional probability with a continuous variable

Nothing changes: $P(A \mid B) = P(A \cap B)/P(B)$, with both probabilities computed as integrals.

HW10 problem 4. A bus arrives at a time X uniform on (0, 30) minutes after 10:00. At 10:20 it has not come; P(you wait at least 5 more minutes)? That is $P(X \geq 25 \mid X > 20) = \frac{P(X \geq 25)}{P(X > 20)} = \frac{5/30}{10/30} = \frac{1}{2}$. Given the bus is late, its arrival is uniform on the remaining 10 minutes.

5.5 The distribution of Y = g(X): the cdf method

> RECIPE. To find the density of $Y = g(X)$:
>
> 1. Find the range of Y.
> 2. Write $F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$, solve the inequality for X, and express it through F_X or an integral of f_X.
> 3. Differentiate: $f_Y(y) = F_Y'(y)$ on the range, 0 outside.
>
> If g is increasing with inverse g^{-1}, this gives $f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}\, g^{-1}(y)$. (Decreasing: take the absolute value.)
> But do the three steps; they never fail, even when g is not one-to-one.

HW10 problem 6(4). X has density $(3/128)\ x^2$ on $[-4, 4]$; $Y = X^2$. Range: $[0, 16]$. For $0 \le y \le 16$,

$$F_Y(y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{3}{128}\ x^2\ dx = \frac{3}{128}\ \cdot\ \frac{2\ y^{3/2}}{3}\ =\ \frac{y^{3/2}}{64}.$$

$$f_Y(y) = \frac{3}{128}\ \sqrt{y} \text{ for } 0 \le y \le 16, \text{ and } 0 \text{ otherwise.}$$

Check: $\int_0^{16} (3/128)\ y^{1/2}\ dy = (3/128)(2/3)(16^{3/2}) = (1/64)(64) = 1$.

Practice set, problem 21. X uniform on $[-0.5, 0.5]$ (density 1 there), $Y = X^3$. Range: $[-1/8, 1/8]$. The cube is increasing, so for y in the range

$$F_Y(y) = P(X \le y^{1/3}) = y^{1/3} + 0.5,\ f_Y(y) = \frac{1}{3}\ y^{-2/3}.$$

At $y = 0.05$: $(1/3)(0.05)^{-2/3} \approx (1/3)(7.368) \approx 2.456$, in $(1.5, 3]$. A density above 1 is normal; Y is squeezed into an interval of length 1/4.

A linear change. If $Y = aX + b$ with $a > 0$, then $F_Y(y) = F_X((y - b)/a)$ and $f_Y(y) = \frac{1}{a}\ f_X(\frac{y - b}{a})$. So if X is uniform on (0, 1), then 30X is uniform on (0, 30), and if Z is standard Normal then $\mu + \sigma Z$ is Normal with mean μ and standard deviation σ (section 13).

5.6 Check yourself

1. f(x) = 2x on [0, 1]. Check it is a density; find P(X > 1/2), F(x), and the median (the m with F(m) = 1/2).

$\int_0^1 2x\ dx = 1$. $P(X > 1/2) = 1 - (1/2)^2 = 3/4$. $F(x) = x^2$ on $[0, 1]$. Median: $m^2 = 1/2$, $m = 1/\sqrt{2} \approx 0.7071$.

2. f(x) = c/x⁴ for x ≥ 10. Find c and P(X > 20). [HW10 P3]

$c = 3000$; $P(X > 20) = (10/20)^3 = 1/8 = 0.125$. (In general $P(X > a) = (10/a)^3$ for $a \ge 10$.)

3. f(x) = c x² on [−4, 4]. Find c, P(|X| = 1/2) and P(|X| > 2). [HW10 P6]

$c = 3/128$; 0; $7/8 = 0.875$.

4. f(x) = ½ e^{−|x|}. Find P(|X| ≤ 2), P(|X| ≤ 2 or X ≤ −1), and P(X is an integer). [HW10 P5]

$1 - e^{-2} \approx 0.8647$; $P(X \le 2) = 1 - \frac{1}{2}e^{-2} \approx 0.9323$; 0.

5. The bus arrives uniformly between 10:00 and 10:30. At 10:20 it has not arrived. P(wait at least 5 more minutes)? [HW10 P4]

$P(X \ge 25 \mid X > 20) = (5/30)/(10/30) = 1/2$.

6. X is uniform on (0, 1) and Y = −ln X. Find P(Y > y) for y > 0 and identify the distribution of Y.

$P(Y > y) = P(-\ln X > y) = P(X < e^{-y}) = e^{-y}$. So $F_Y(y) = 1 - e^{-y}$: Y is Exponential with rate 1 (section 10). This is how computers simulate exponential variables.

7. X has density $3x^2$ on $[0, 1]$; $Y = X^2$. Find f_Y.

Range $[0, 1]$. $F_Y(y) = P(X \le \sqrt{y}) = (\sqrt{y})^3 = y^{3/2}$, so $f_Y(y) = (3/2)\sqrt{y}$ on $[0, 1]$.

8. X uniform on $[-0.5, 0.5]$, $Y = X^3$. Find $f_Y(0.05)$. [Practice 21]

$f_Y(y) = (1/3)y^{-2/3}$ on $[-1/8, 1/8]$; at 0.05 it is ≈ 2.456.

Deck 6 — 6. Expectation, variance and functions of a random variable

The idea. The expected value is the long-run average of a random variable: what you would get per trial if you repeated the experiment forever. The variance measures how spread out the values are around that average. Both are computed from the pmf (sum) or the pdf (integral), and both obey a few algebra rules that you will use on every exam question.

6.1 Expected value of a discrete random variable

Definition. If X takes values x_i with probabilities $p(x_i)$,

$$E(X) = \Sigma_i \, x_i \, p(x_i).$$

A weighted average of the values, weighted by their probabilities. Also written EX, μ, or "the mean".

Why this is the long-run average (slides). Repeat the experiment n times. The average of the results is $\Sigma_k \, x_k \cdot$ (number of times x_k occurred)/n, and as n grows each frequency (count)/n settles down to $p(x_k)$. So the average tends to $\Sigma \, x_k \, p(x_k)$. HW7 problem 5 phrases it this way: among a huge number of candidates like Brad, the average number of interviews would be close to E(X).

- **A fair die.** $E(X) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$. The mean need not be a value X can take; do not be confused by the word "expected".
- **An indicator.** I = 1 if A occurs, 0 otherwise. $E(I) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A)$. Expectation of an indicator is a probability; this will matter a lot in section 11.
- **Brad's interviews (HW7 problem 5).** X = 0, 1, 2, 3 with probabilities 0.1, 0.2, 0.6, 0.1: $E(X) = 0 + 0.2 + 1.2 + 0.3 = 1.7$.
- **HW7 problem 3(b).** Values −1.3, 1.2, 2.7, 3.3, 9.8 with probabilities 0.1, 0.25, 0.35, 0.1, 0.2: $E(X) = -0.13 + 0.30 + 0.945 + 0.33 + 1.96 = 3.405$.
- **From a cdf (practice set, problem 8).** Read the pmf off the jumps first (section 4.3), then E(X) = 0.25.

6.2 Expectation of a function of X

Plug-in rule. If g is any function, g(X) is a random variable and

$$E[g(X)] = \Sigma_i \, g(x_i) \, p(x_i).$$

You do not need the pmf of g(X): keep the probabilities of X and plug each value into g. $E(X^n)$ is the n-th moment.

Example (slides). X = −1, 0, 1 with probabilities 0.2, 0.5, 0.3 and $Y = X^2$. The long way: Y = 0 with probability 0.5 and Y = 1 with probability 0.5, so E(Y) = 0.5. The plug-in way: $0.2(-1)^2 + 0.5(0)^2 + 0.3(1)^2 = 0.5$. Same number, and the second way needs no new pmf.

HW7 problem 3(b), continued. $E(e^X) = 0.1e^{-1.3} + 0.25e^{1.2} + 0.35e^{2.7} + 0.1e^{3.3} + 0.2e^{9.8} \approx 0.027 + 0.830 + 5.206 + 2.711 + 3606.75 \approx 3615.5$. One large value dominates; that is typical of e^X, and it is the reason MGFs (section 7) are so informative.

HW7 problem 4 (stocking a product). Each unit sold earns \$5; each unsold unit loses \$15. Demand X = 2, 3, 4 with probabilities 0.1, 0.2, 0.7. If the store stocks s units, profit $= 5 \cdot \min(X, s) - 15 \cdot \max(s - X, 0)$. Compute the expected profit for each s:

- s = 2: every demand buys both units, profit 10 always, E P(2) = 10.
- s = 3: demand 2 gives 10 − 15 = −5 (probability 0.1); demand 3 or 4 gives 15 (probability 0.9). E P(3) = −0.5 + 13.5 = 13.

- s = 4: demand 2 gives 10 − 30 = −20 (0.1); demand 3 gives 15 − 15 = 0 (0.2); demand 4 gives 20 (0.7). E P(4) = −2 + 0 + 14 = 12.

So E P(2) < E P(4) < E P(3): option (B). Stock 3. The method is the whole point: list the values of the function for each outcome, then weight by the probabilities.

E[g(X)] is not g(E X). $E(X^2) \neq (E\ X)^2$ in general (for the die, 91/6 ≈ 15.17 versus 12.25), and E(1/X) ≠ 1/E(X). The only functions that commute with E are linear ones, next.

6.3 Linearity

E(aX + b) = a E(X) + b. Proof: $\Sigma\ (a\ x_i + b)\ p(x_i) = a\ \Sigma\ x_i\ p(x_i) + b\ \Sigma\ p(x_i) = a\ E(X) + b$. (Section 11 extends this: E(X + Y) = E(X) + E(Y) for any two random variables, independent or not.)

HW7 problem 5(8). Each interview costs 10 hours of preparation and the applications took 20 hours, so Y = 10X + 20 and E(Y) = 10(1.7) + 20 = 37 hours.

6.4 Variance and standard deviation

The mean says nothing about spread: W = 0, Y = ±1 (each with probability 1/2) and Z = ±100 all have mean 0. The variance distinguishes them.

Variance. If X has mean μ,

$$\mathrm{Var}(X) = E[(X - \mu)^2] = E(X^2) - (E\ X)^2.$$

The standard deviation is SD(X) = $\sqrt{\mathrm{Var}(X)}$, in the same units as X. The second form is the one you compute with: find $E(X^2)$ by the plug-in rule, subtract the square of the mean.

Why the two forms agree (slides): $E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = E(X^2) - 2\mu\ E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$, using linearity and the fact that μ is a constant.

- **A fair die.** $E(X^2) = (1 + 4 + 9 + 16 + 25 + 36)/6 = 91/6$. $\mathrm{Var}(X) = 91/6 - (7/2)^2 = 91/6 - 49/4 = 35/12 \approx 2.917$; SD ≈ 1.708.
- **Brad (HW7 problem 5).** $E(X^2) = 0 + 0.2 + 2.4 + 0.9 = 3.5$; $\mathrm{Var}(X) = 3.5 - 1.7^2 = 3.5 - 2.89 = 0.61$.
- **W, Y, Z above.** Var(W) = 0, Var(Y) = 1, Var(Z) = 10,000.

Var(aX + b) = a^2 Var(X). Shifting by b does nothing to the spread; scaling by a scales the standard deviation by |a| and the variance by a^2. Proof (slides): aX + b − E(aX + b) = a(X − μ), so $E[(a(X - \mu))^2] = a^2\ E[(X - \mu)^2]$.

HW7 problem 5(9). Var(Y) = Var(10X + 20) = 100 · 0.61 = 61 hours^2; SD(Y) = $\sqrt{61} \approx 7.81$ hours.

Common slips. Var(aX + b) has no b in it and has a^2, not a. Variance is never negative; if you get a negative number, you subtracted $(E\ X)^2$ from the wrong thing. And SD, not variance, is in the units of X: the practice set asks for SD(X) in problem 9 and for Var(C) in problem 10.

6.5 The continuous versions

SUMS BECOME INTEGRALS. If X has density f,

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx,\quad E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx,\quad Var(X) = E(X^2) - (E\, X)^2 \text{ with } E(X^2) = \int x^2\, f(x)\, dx.$$

Linearity and $Var(aX + b) = a^2 Var(X)$ hold exactly as before (slides: pull the constants out of the integral).

Uniform on (0, 1). $E(X) = \int_0^1 x \cdot 1\, dx = 1/2$. $E(X^2) = \int_0^1 x^2\, dx = 1/3$. $Var(X) = 1/3 - 1/4 = 1/12$.

The chocolate bar (slides). A 1-foot bar is broken at a uniform point U. Amy gets the piece containing a fixed point p. If $U < p$ her piece is $[U, 1]$, length $1 - U$; if $U > p$ it is $[0, U]$, length U. So the length $L = g(U)$ and

$$E(L) = \int_0^p (1 - u)\, du + \int_p^1 u\, du = (p - \frac{p^2}{2}) + (\frac{1}{2} - \frac{p^2}{2}) = \frac{1}{2} + p - p^2.$$

At $p = 1/2$ this is 3/4: the piece containing the middle is on average three quarters of the bar, because whichever side the break lands on, Amy gets the bigger side plus a bit. The plug-in rule with a piecewise g is exactly this: split the integral where the formula for g changes.

HW10 problem 3, continued. $f(x) = 3000/x^4$ for $x \geq 10$.

- $E(X) = \int_{10}^{\infty} 3000\, x^{-3}\, dx = 3000 \cdot \frac{10^{-2}}{2} = 15$.
- $E(X^2) = \int_{10}^{\infty} 3000\, x^{-2}\, dx = 3000 \cdot 10^{-1} = 300$, so $Var(X) = 300 - 225 = 75$.
- $E(2X^2 + 3X) = 2(300) + 3(15) = 645$ by linearity.

(With c/x^3 instead of c/x^4 the variance integral would diverge: heavy-tailed variables can have a mean but infinite variance. The homework chose the exponent 4 so that everything exists.)

6.6 Check yourself

1. Brad: X = 0, 1, 2, 3 with probabilities 0.1, 0.2, 0.6, 0.1. Find E(X), $E(X^2)$, Var(X); then E(Y) and Var(Y) for Y = 10X + 20. [HW7 P5]

E(X) = 1.7; $E(X^2) = 3.5$; Var(X) = 0.61; E(Y) = 37; Var(Y) = 61.

2. Demand X = 2, 3, 4 with probabilities 0.1, 0.2, 0.7; \$5 profit per unit sold, \$15 loss per unsold unit. Order the expected profits for stocking 2, 3, 4 units. [HW7 P4]

E P(2) = 10, E P(3) = 13, E P(4) = 12; E P(2) < E P(4) < E P(3).

3. X = −1, 0, 1 with probabilities 0.2, 0.5, 0.3. Find E(X), $E(X^2)$ and Var(X).

$E(X) = -0.2 + 0.3 = 0.1$; $E(X^2) = 0.5$; $Var(X) = 0.5 - 0.01 = 0.49$.

4. You pay \$2 to roll a fair die and receive the face value in dollars. Expected net gain? Variance of the net gain?

$E(X - 2) = 3.5 - 2 = 1.5$. $Var(X - 2) = Var(X) = 35/12 \approx 2.917$.

5. $f(x) = 3000/x^4$ for $x \geq 10$. Find $E(X)$, $Var(X)$ and $E(2X^2 + 3X)$. [HW10 P3]

15; 75; 645.

6. In the chocolate-bar problem with $p = 1/4$, what is the expected length of Amy's piece?

$1/2 + 1/4 - 1/16 = 11/16 = 0.6875$.

7. X uniform on (0, 1). Find $E(X^3)$ and $Var(3X - 2)$.

$E(X^3) = \int_0^1 x^3\,dx = 1/4$. $Var(3X - 2) = 9 \cdot (1/12) = 3/4$.

8. Values −1.3, 1.2, 2.7, 3.3, 9.8 with probabilities 0.1, 0.25, 0.35, 0.1, 0.2. Find $E(X)$ and $E(e^X)$. [HW7 P3]

3.405 and ≈ 3615.5.

9. $f(x) = \tfrac{1}{2}\, e^{-|x|}$ (Laplace). Find $E(X)$ and $Var(X)$.

$E(X) = 0$ by symmetry. $Var(X) = E(X^2) = 2\int_0^\infty x^2 \cdot \tfrac{1}{2}\, e^{-x}\,dx = \int_0^\infty x^2\, e^{-x}\,dx = 2! = 2$.

Deck 6 7. Moment generating functions

The idea. The moment generating function packs every moment $E(X)$, $E(X^2)$, $E(X^3)$, … of a random variable into one function of t. Differentiate it at 0 and the moments fall out. More importantly, the MGF is the "DNA" of a distribution: two random variables with the same MGF have the same distribution, so you can recognize a Binomial or a Poisson from its MGF, and you can find the distribution of a sum of independent variables by multiplying MGFs.

7.1 Definition

MGF.

$$M_X(t) = E(e^{tX}) = \Sigma_i\, e^{t\, x_i}\, p(x_i) \text{ (discrete) or } \int e^{tx}\, f(x)\, dx \text{ (continuous), for the t where this is finite.}$$

Always $M_X(0) = E(1) = 1$. It is just the plug-in rule with $g(x) = e^{tx}$, where t is a bookkeeping variable.

HW7 problem 5(3). X = 0, 1, 2, 3 with probabilities 0.1, 0.2, 0.6, 0.1. $M_X(t) = 0.1 + 0.2e^t + 0.6e^{2t} + 0.1e^{3t}$. For a finite table the MGF is just "probability times $e^{t \cdot \text{value}}$, added up".

7.2 Derivatives at 0 give the moments

Moments from the MGF.

$$M_X'(0) = E(X),\ M_X''(0) = E(X^2), \text{ and in general the n-th derivative at 0 is } E(X^n).$$

So $\text{Var}(X) = M''(0) - M'(0)^2$.

Why (slides): differentiating e^{tX} with respect to t gives $X\, e^{tX}$, so $M'(t) = E(X\, e^{tX})$ and $M'(0) = E(X\, e^0) = E(X)$. Differentiate again: $M''(t) = E(X^2\, e^{tX})$, $M''(0) = E(X^2)$. (Moving the derivative inside the sum or integral is allowed whenever the MGF exists near 0.)

HW7 problem 5(4)–(7). $M'(t) = 0.2e^t + 1.2e^{2t} + 0.3e^{3t}$, so $M'(0) = 0.2 + 1.2 + 0.3 = 1.7 = E(X)$. $M''(t) = 0.2e^t + 2.4e^{2t} + 0.9e^{3t}$, $M''(0) = 3.5 = E(X^2)$, $\text{Var}(X) = 3.5 - 2.89 = 0.61$. Exactly the numbers from section 6, which is the point: the MGF is a machine that produces them.

7.3 The MGF determines the distribution

Uniqueness. If $M_X(t) = M_Y(t)$ for all t in an interval around 0, then X and Y have the same distribution. Consequently, if you can write an MGF as "Σ (probability) $\cdot\ e^{t \cdot \text{value}}$", you can read the pmf straight off it.

HW7 problem 10 (Sherlock's scrap of paper). $M_X(t) = \frac{3}{4}\, e^{2t} + \frac{1}{8}\, e^t + \frac{1}{8}$. Match the pattern: X = 2 with probability 3/4, X = 1 with probability 1/8, X = 0 with probability 1/8 (the constant is $e^{0 \cdot t}$). Then $E(X) = 2(3/4) + 1(1/8) = 1.625$, $E(X^2) = 4(3/4) + 1(1/8) = 3.125$, $\text{Var}(X) = 3.125 - 1.625^2 = 3.125 - 2.6406 = 0.4844 \approx 0.48$. Differentiating twice gives the same thing; reading off the pmf is faster.

HW9 problem 8 (Bert and Ernie). $M_X(t) = (1/3 + (2/3)e^t)^5$ is the Binomial MGF $(q + pe^t)^n$ with $n = 5$, $p = 2/3$ (section 8), and $M_Y(t) = e^{2(e^t - 1)}$ is the Poisson MGF with $\lambda = 2$ (section 9). Recognizing them gives $E(X) = 10/3$, $\mathrm{Var}(X) = 10/9$, $E(Y) = 2$, $\mathrm{Var}(Y) = 2$ instantly; the rest of that problem ($Z = XY$) is finished in section 12.

How it is asked. "X has MGF … . Compute Var(X)." Either recognize a famous MGF (section 17 has the table) or expand it into probabilities times exponentials. Only differentiate as a last resort; it is slow and error-prone with products.

7.4 MGFs of linear transforms and of sums

Two rules.

$$M_{aX+b}(t) = e^{bt}\, M_X(at), \text{ because } E(e^{t(aX+b)}) = e^{bt}\, E(e^{(at)X}).$$

If X and Y are independent: $M_{X+Y}(t) = M_X(t)\, M_Y(t)$, and more generally $M_{aX+bY}(t) = M_X(at)\, M_Y(bt)$.

The second rule uses $E(e^{tX}\, e^{tY}) = E(e^{tX})\, E(e^{tY})$, which is true for independent variables (section 12) and false in general.

Practice set, problem 15. X, Y independent; the MGF of $3X - 2Y$ is $M_X(3t)\, M_Y(-2t)$: option (C). Options with powers like $(M_X(t))^3$ confuse "3X" with "$X_1 + X_2 + X_3$" (three independent copies would give $M_X(t)^3$); options with sums of MGFs are never right.

HW10 problem 8. (1) $M_{X-2Y}(t) = M_X(t)\, M_Y(-2t)$. (2) Without independence, no: the MGF of $X - 2Y$ depends on the joint distribution. If $Y = X$ (perfectly dependent) then $X - 2Y = -X$ has MGF $M_X(-t)$, while an independent copy gives $M_X(t)M_X(-2t)$; same marginals, different answers.

HW9 problem 5. $X \sim \text{Binomial}(4, 0.3)$ and $Y \sim \text{Poisson}(2)$ independent; $E(e^{X+Y}) = M_{X+Y}(1) = M_X(1)\, M_Y(1) = (0.7 + 0.3e)^4 \cdot e^{2(e-1)} \approx 5.2748 \times 31.080 \approx 163.94$. An MGF evaluated at $t = 1$ is just $E(e^X)$.

Sums of famous variables. Multiply the MGFs and recognize the result: Binomial(n, p) + Binomial(m, p) = Binomial(n + m, p); Poisson(λ) + Poisson(μ) = Poisson(λ + μ); Normal + Normal = Normal with the means and variances added. These are used in sections 9, 12 and 13.

7.5 MGFs you should recognize

Distribution	$M_X(t)$	Section
Bernoulli(p)	$1 - p + pe^t$	8
Binomial(n, p)	$(1 - p + pe^t)^n$	8
Poisson(λ)	$e^{\lambda(e^t - 1)}$	9
Geometric(p)	$\dfrac{pe^t}{1 - (1-p)e^t}$ for $t < -\ln(1-p)$	11
Uniform(a, b)	$\dfrac{e^{tb} - e^{ta}}{t(b-a)}$	10

Exponential(λ)	$\frac{\lambda}{\lambda - t}$ for $t < \lambda$ (infinite otherwise)	10
Normal(μ, σ^2)	$e^{\mu t + \sigma^2 t^2/2}$	13

7.6 Check yourself

1. X = 0, 1, 2, 3 with probabilities 0.1, 0.2, 0.6, 0.1. Write $M_X(t)$, $M_X'(t)$, and use them to get E(X), $E(X^2)$, Var(X). [HW7 P5]

$M(t) = 0.1 + 0.2e^t + 0.6e^{2t} + 0.1e^{3t}$; $M'(t) = 0.2e^t + 1.2e^{2t} + 0.3e^{3t}$; $M'(0) = 1.7$; $M''(0) = 3.5$; Var = 0.61.

2. $M_X(t) = (3/4)e^{2t} + (1/8)e^t + 1/8$. Find Var(X). [HW7 P10]

X = 2, 1, 0 with probabilities 3/4, 1/8, 1/8. $E(X) = 1.625$, $E(X^2) = 3.125$, $Var(X) \approx 0.48$.

3. Identify the distribution, mean and variance: (a) $M(t) = (0.4 + 0.6e^t)^{10}$; (b) $M(t) = e^{3(e^t - 1)}$; (c) $M(t) = 2/(2 - t)$.

(a) Binomial(10, 0.6): mean 6, variance 2.4. (b) Poisson(3): mean 3, variance 3. (c) Exponential with rate 2: mean 1/2, variance 1/4.

4. X, Y independent with MGFs M_X, M_Y. MGF of 3X − 2Y? Of X − 2Y? Can you find the MGF of X − 2Y without independence? [Practice 15, HW10 P8]

$M_X(3t)M_Y(-2t)$; $M_X(t)M_Y(-2t)$; no, the joint distribution is needed.

5. X ~ Binomial(4, 0.3), Y ~ Poisson(2), independent. $E(e^{X+Y})$? [HW9 P5]

$M_X(1)M_Y(1) = (0.7 + 0.3e)^4 \, e^{2(e-1)} \approx 163.94$.

6. $M_X(t) = e^{t^2/2}$. Find E(X) and $E(X^2)$.

$M'(t) = t\, e^{t^2/2}$, so $E(X) = 0$. $M''(t) = (1 + t^2)\, e^{t^2/2}$, so $E(X^2) = 1$. (This is the standard Normal.)

7. If $M_X(t) = \frac{1}{2}(1 + e^t)$, what is the MGF of 2X + 1, and what are its values and probabilities?

$M_{2X+1}(t) = e^t M_X(2t) = \frac{1}{2}(e^t + e^{3t})$: the values 1 and 3, each with probability 1/2 (X is a fair coin, 0 or 1).

8. X_1, X_2, X_3 are independent Poisson(1), Poisson(2), Poisson(4). What is the distribution of $X_1 + X_2 + X_3$?

Product of the MGFs is $e^{(1+2+4)(e^t - 1)}$: Poisson(7).

Deck 6 8. Bernoulli and Binomial random variables

The idea. The slides call the famous distributions the "X-men" and study each one with the same workflow: discrete or continuous? origin story? properties (mean, variance, MGF)? applications? The Binomial is the first: it counts successes in a fixed number of independent, identical yes/no trials. If you can say "n independent trials, each a success with probability p, and I am counting successes", you have a Binomial.

8.1 Bernoulli trials

Bernoulli(p). A trial with two outcomes, success (probability p) and failure (probability $1 - p$). The Bernoulli random variable X is 1 for success and 0 for failure. $E(X) = p$, $E(X^2) = p$, so $\mathrm{Var}(X) = p - p^2 = p(1 - p)$, and $M_X(t) = (1 - p) + pe^t$. An indicator I_A is a Bernoulli(P(A)).

8.2 The Binomial distribution

Binomial(n, p). X = number of successes in n independent Bernoulli(p) trials. Values 0, 1, …, n and

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i},\ i = 0, 1, \ldots, n.$$

Notation $X \sim B(n, p)$ or Bin(n, p); Bernoulli is B(1, p). A fair coin tossed n times gives B(n, 1/2) heads.

Where the formula comes from (slides, 3 trials, p = 1/4). $P(X = 0) = P(FFF) = (3/4)^3$. $P(X = 1) = P(SFF) + P(FSF) + P(FFS) = 3 \cdot (1/4)(3/4)^2$: each sequence with exactly one success has the same probability, and there are $\binom{3}{1} = 3$ positions for the success. In general each sequence with i successes has probability $p^i(1 - p)^{n-i}$ (independence), and there are $\binom{n}{i}$ such sequences. The probabilities add to 1 by the binomial theorem: $\Sigma \binom{n}{i} p^i(1 - p)^{n-i} = (p + 1 - p)^n = 1$.

8.3 Mean, variance and MGF

Binomial facts. If $X \sim B(n, p)$ and $q = 1 - p$:

$$E(X) = np,\ \mathrm{Var}(X) = np(1 - p),\ \mathrm{SD}(X) = \sqrt{(np(1 - p))},\ M_X(t) = (q + pe^t)^n.$$

The MGF (slides). $M_X(t) = \Sigma_k\, e^{tk} \binom{n}{k} p^k q^{n-k} = \Sigma_k \binom{n}{k} (pe^t)^k q^{n-k} = (q + pe^t)^n$, by the binomial theorem with $x = q$ and $y = pe^t$. Then $M'(t) = n(q + pe^t)^{n-1} pe^t$, so $E(X) = M'(0) = np$. One more derivative (product rule) gives $M''(0) = np + n(n - 1)p^2 = E(X^2)$, hence $\mathrm{Var}(X) = np + n^2p^2 - np^2 - n^2p^2 = np(1 - p)$. (Section 12 gives the quick proof: X is a sum of n independent Bernoullis, so means and variances add.)

The variance is largest at $p = 1/2$ and shrinks to 0 at $p = 0$ or 1: a nearly certain outcome has little spread.

8.4 Worked examples

HW7 problem 6: headlights. Each of 50 cars forgets its lights with probability 0.04, independently. $X \sim B(50, 0.04)$. $P(X = 2) = \binom{50}{2}(0.04)^2(0.96)^{48} = 1225 \cdot 0.0016 \cdot 0.1409 \approx 0.2762$.

HW7 problem 9: Brad's exam. 80 questions, 5 choices each. Brad knows the answer with probability 0.75; otherwise he guesses. Per question, P(correct) = 0.75 + 0.25(1/5) = 0.80 (total probability). Questions are independent, so $X \sim B(80, 0.8)$: $E(X) = 64$, $\mathrm{Var}(X) = 80(0.8)(0.2) = 12.8$.

HW7 problem 8: flu vaccines. 300 employees all get the GS vaccine, which is 80% effective, so each still gets the flu with probability 0.2, independently: X ~ B(300, 0.2) employees fall ill. Cost C = 300 · \$40 + \$300 · X = 12,000 + 300X. By linearity and the scaling rule, E(C) = 12,000 + 300(60) = \$30,000 and Var(C) = 300^2 · Var(X) = 90,000 · 300(0.2)(0.8) = 90,000 · 48 = 4,320,000.

Practice set, problems 9–10 (same story, 40 employees, 70% effective, \$70 per vaccine, \$1000 per sick employee). X ~ B(40, 0.3): E(X) = 12, SD(X) = $\sqrt{(40 \cdot 0.3 \cdot 0.7)} = \sqrt{8.4} \approx 2.90$, option (D). C = 2800 + 1000X, so Var(C) = 1000^2 · 8.4 = 8,400,000, in [10,000, ∞), option (E). The fixed vaccine cost contributes nothing to the variance.

Motco's boxes (slides): a Binomial inside a Binomial. A box holds 10 packets; each packet has raisins with probability 0.05, independently. A box is returned if more than one packet has raisins. Eva buys 5 boxes; P(she returns exactly one)?

1. Inside one box: Y = number of raisin packets ~ B(10, 0.05). The box is returned with probability p = P(Y ≥ 2) = 1 − P(Y = 0) − P(Y = 1) = $1 - 0.95^{10} - 10(0.05)(0.95)^9 = 1 - 0.5987 - 0.3151 \approx 0.0861$.
2. Across boxes: each box is returned with probability p independently, so X = number returned ~ B(5, p) and P(X = 1) = $\binom{5}{1}$ p $(1 - p)^4 = 5(0.0861)(0.9139)^4 \approx 0.3004$.

HW7 problem 7: pooled testing. 1000 employees, each infected with probability 0.05; samples are pooled 10 at a time, so there are 100 pools. A pool tests positive if at least one of its 10 members is infected: probability $1 - 0.95^{10} \approx 0.4013$. A negative pool costs 1 test; a positive pool costs 1 + 10 tests (everyone is retested). Expected tests per pool = 1 + 10(0.4013) = 5.013; expected total = 100 × 5.013 ≈ 501.3, about half of testing everyone individually.

Airplane engines (slides, from the textbook). Engines fail independently with probability 1 − p; a plane needs a majority of its engines. A 3-engine plane succeeds with probability $3p^2(1 - p) + p^3$; a 5-engine plane with $\binom{5}{3}p^3(1 - p)^2 + \binom{5}{4}p^4(1 - p) + p^5$. At p = 0.6 these are 0.648 and 0.683; at p = 0.4 they are 0.352 and 0.317; at p = 1/2 both are 1/2. The 5-engine plane is preferable exactly when p > 1/2: redundancy only helps when the components are reliable to begin with. (The slides quote Downer's article on redundancy in complex systems for the engineering context.)

8.5 Recognizing a Binomial (and when it is not one)

Story	Binomial?
Fixed number n of independent trials, same p each time, count the successes	Yes: B(n, p)
Draws without replacement from a finite population (aces in a 5-card hand)	No: trials are dependent (use counting; section 11.2 for the joint version)
Number of trials until the first success	No: Geometric (section 11.1)
Rare events, large n, small p	Binomial, but approximate it by a Poisson (section 9)
Trials with different success probabilities	Not Binomial, but means still add (section 11.5)

How it is asked. "X has a Binomial distribution. What are n and p?" Then "compute P(X = 2)", "E(X) and SD(X)", or a cost that is a linear function of X. Write n and p down explicitly before computing anything; p is often itself the result of a total-probability step, as in Brad's exam.

8.6 Check yourself

1. 50 cars, each forgets its headlights with probability 0.04. P(exactly 2 forget)? [HW7 P6]

$\binom{50}{2}(0.04)^2(0.96)^{48} \approx 0.2762$.

2. 80 questions, 5 choices; Brad knows each answer with probability 0.75 and guesses otherwise. E(X) and Var(X) for X = number correct? [HW7 P9]

p = 0.75 + 0.25/5 = 0.8; X ~ B(80, 0.8); E = 64, Var = 12.8.

3. 40 vaccinated employees, vaccine 70% effective, $70 per vaccine, $1000 per sick employee. E(X), SD(X) for X = number sick, and Var(C) for the total cost C? [Practice 9–10]

X ~ B(40, 0.3): E = 12, SD = $\sqrt{8.4} \approx 2.90$. C = 2800 + 1000X: Var(C) = 8,400,000.

4. Motco: 10 packets per box, raisins with probability 0.05 each, box returned if more than one packet has raisins. Eva buys 5 boxes; P(returns exactly one)?

$p = 1 - 0.95^{10} - 10(0.05)(0.95)^9 \approx 0.0861$; $5p(1-p)^4 \approx 0.3004$.

5. 1000 employees, infection probability 0.05, pools of 10. Expected number of tests per day? [HW7 P7]

$100 \text{ pools} \times [1 + 10(1 - 0.95^{10})] \approx 501.3$.

6. 300 employees vaccinated with an 80% effective vaccine at $40 each; a sick employee costs $300. E(C) and Var(C)? [HW7 P8]

X ~ B(300, 0.2); C = 12,000 + 300X; E(C) = 30,000; Var(C) = 90,000 · 48 = 4,320,000.

7. A fair coin is tossed 10 times. P(exactly 5 heads)? P(at least 8 heads)?

$252/1024 \approx 0.2461$; $(45 + 10 + 1)/1024 = 56/1024 \approx 0.0547$.

8. A Binomial X has E(X) = 6 and Var(X) = 2.4. Find n and p.

Var/E = 1 − p = 0.4, so p = 0.6 and n = 6/0.6 = 10.

9. Each trial succeeds with probability 0.1. How many independent trials are needed so that P(at least one success) ≥ 0.95?

$1 - 0.9^n \geq 0.95 \Leftrightarrow 0.9^n \leq 0.05 \Leftrightarrow n \geq \ln(0.05)/\ln(0.9) \approx 28.4$, so n = 29.

10. For the 3- and 5-engine planes with p = 0.6, which is safer?

3 engines: $3(0.36)(0.4) + 0.216 = 0.648$. 5 engines: $10(0.216)(0.16) + 5(0.1296)(0.4) + 0.0778 \approx 0.683$. The 5-engine plane, because $p > 1/2$.

Deck 6 9. Poisson random variables and the Poisson process

The idea. The Poisson distribution counts rare events: misprints on a page, bombs falling on a quarter-square-kilometer of London, goals by Gretzky in a game, customers at a post office. It is what a Binomial becomes when the number of trials is huge and the success probability tiny, and it is the distribution of the number of events in a time interval when events occur "at random" at a constant rate. One parameter, λ, is both its mean and its variance.

9.1 Definition and basic properties

Poisson(λ). X takes values 0, 1, 2, … with

$$P(X = i) = e^{-\lambda}\,\frac{\lambda^i}{i!},\ i = 0, 1, 2, \ldots, \text{ for some } \lambda > 0.$$

$$E(X) = \lambda,\ \mathrm{Var}(X) = \lambda,\ M_X(t) = e^{\lambda(e^t - 1)}.$$

Why $e^{-\lambda}$ (slides). The probabilities must add to 1: $\Sigma_i\, e^{-\lambda}\,\lambda^i/i! = e^{-\lambda}\,\Sigma_i\,\lambda^i/i! = e^{-\lambda}\,e^{\lambda} = 1$, by the Taylor series of e^{λ}. The $e^{-\lambda}$ is exactly the normalizing constant.

The MGF (slides). $M(t) = \Sigma\, e^{ti}\, e^{-\lambda}\,\lambda^i/i! = e^{-\lambda}\,\Sigma\,(\lambda e^t)^i/i! = e^{-\lambda}\,e^{\lambda e^t} = e^{\lambda(e^t - 1)}$. Then $M'(t) = \lambda e^t\, e^{\lambda(e^t - 1)}$, so $E(X) = M'(0) = \lambda$; $M''(t) = (\lambda e^t)^2\, e^{\lambda(e^t - 1)} + \lambda e^t\, e^{\lambda(e^t - 1)}$, so $E(X^2) = \lambda^2 + \lambda$ and $\mathrm{Var}(X) = \lambda$.

Fast probabilities. $P(X = 0) = e^{-\lambda}$, $P(X = 1) = \lambda e^{-\lambda}$, $P(X = 2) = \lambda^2 e^{-\lambda}/2$. "At least one" $= 1 - e^{-\lambda}$. "More than two" $= 1 - e^{-\lambda}(1 + \lambda + \lambda^2/2)$. You will compute these dozens of times; $e^{-\lambda}$ factors out of everything.

9.2 The Poisson as a limit of the Binomial

Poisson approximation. If $Y \sim B(n, p)$ with n large, p small and np moderate, then Y is approximately Poisson(λ) with $\lambda = np$.

Why (slides). Put $p = \lambda/n$. Then $P(Y = k) = \binom{n}{k}\,(\lambda/n)^k\,(1 - \lambda/n)^{n-k} = \frac{n(n-1)\cdot\cdots\cdot(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\cdot(1 - \lambda/n)^n\cdot(1 - \lambda/n)^{-k}$. As $n \to \infty$ the first factor $\to 1$, $(1 - \lambda/n)^n \to e^{-\lambda}$, and the last factor $\to 1$, leaving $e^{-\lambda}\lambda^k/k!$. The mean $np = \lambda$ and the variance $np(1 - p) \approx \lambda$ match too.

The Poisson paradigm. n events with small probabilities $p_1, \ldots, p_n$, independent or only "weakly dependent": the number that occur is approximately Poisson with mean $\Sigma\, p_i$. This is why birthday-type counts (which are not quite independent) are still Poisson to excellent accuracy.

Evidence from the slides. 537 bombs fell on 576 quarter-square-kilometer squares of South London, $\lambda = 537/576 = 0.93$ per square. Predicted number of squares with 0, 1, 2, 3, 4, 5+ bombs: 226.7, 211.4, 98.5, 30.6, 7.1, 1.6. Observed: 229, 211, 93, 35, 7, 1. Gretzky's points per game ($\lambda = 1669/696 \approx 2.4$) fit a Poisson just as well. Many tiny chances, added up, give a Poisson.

Brad's dorm (slides). 180 other students; X = number sharing Brad's birthday. Exactly, $X \sim B(180, 1/365)$; approximately, Poisson with $\lambda = 180/365 \approx 0.4932$.

- (a) $P(X \geq 1) \approx 1 - e^{-0.4932} \approx 0.3893$. Exact: $1 - (364/365)^{180} \approx 0.3897$.

- (b) $P(X = 1) \approx 0.4932\, e^{-0.4932} \approx 0.3012$. Exact: $180(1/365)(364/365)^{179} \approx 0.3018$.
- (c) $P(X \geq 2) \approx 1 - e^{-0.4932}(1 + 0.4932) \approx 0.0881$.

HW8 problem 1: yelling runners. Brad passes n runners; each yells with probability 0.03.

(1) $n = 100$: $X \approx \text{Poisson}(3)$. $P(X > 2) = 1 - e^{-3}(1 + 3 + 4.5) = 1 - 8.5e^{-3} \approx 0.5768$. (Exact Binomial: 0.5802.)

(2) $P(X \geq 1) = 1 - e^{-0.03n} \geq 0.95 \Leftrightarrow e^{-0.03n} \leq 0.05 \Leftrightarrow n \geq \ln(20)/0.03 \approx 99.86$, so $n = 100$ runners.

Practice set, problem 12. 200 runners, probability 0.01: $X \approx \text{Poisson}(2)$, $P(X > 2) = 1 - e^{-2}(1 + 2 + 2) = 1 - 5e^{-2} \approx 0.3233$, in (0.25, 0.40].

9.3 More worked examples

HW8 problems 2–3. $P(X = 1 \mid X \leq 1) = \dfrac{\lambda e^{-\lambda}}{e^{-\lambda} + \lambda e^{-\lambda}} = \dfrac{\lambda}{1 + \lambda} = 0.5$, so $\lambda = 1$. Conditional probability with a Poisson: both numerator and denominator are sums of pmf terms, and $e^{-\lambda}$ cancels.

HW8 problem 4: does the drug work? (Bayes with a Poisson likelihood). Colds per year are Poisson(5); the drug lowers λ to 3 for 75% of people and does nothing for 25%. A user has 2 colds. Let B = the drug is beneficial for him.

$$P(B \mid 2 \text{ colds}) = \frac{0.75 \cdot e^{-3}3^2/2!}{0.75 \cdot e^{-3}3^2/2! + 0.25 \cdot e^{-5}5^2/2!} = \frac{0.75(0.2240)}{0.75(0.2240) + 0.25(0.0842)} = \frac{0.1680}{0.1891} \approx 0.8886.$$

HW8 problem 5: Clark and Diana. 400 questions; Clark errs with probability 0.0075, Diana 0.003. $X_C \approx \text{Poisson}(3)$, $X_D \approx \text{Poisson}(1.2)$. The sum of independent Poissons is Poisson with the summed mean (multiply the MGFs: $e^{3(e^t-1)}$ $e^{1.2(e^t-1)} = e^{4.2(e^t-1)}$), so $X_C + X_D \approx \text{Poisson}(4.2)$ and $P(X_C + X_D = 2) = e^{-4.2}(4.2)^2/2 \approx 0.1323$.

Practice set, problem 17. $X \sim \text{Poisson}(1)$, $Y \sim \text{Poisson}(2)$, independent. $P(\max(X, Y) \leq 1) = P(X \leq 1 \text{ and } Y \leq 1) = P(X \leq 1)\, P(Y \leq 1) = (e^{-1} + e^{-1})(e^{-2} + 2e^{-2}) = 2e^{-1} \cdot 3e^{-2} = 6e^{-3} \approx 0.2987$, in [0, 0.5). (Section 12.4 has the general max/min method.)

Two buses (slides). Bus 1 carries Poisson(4) passengers, bus 2 Poisson(5); you take bus 1 with probability 1/3. P(at least 2 passengers) $= \frac{1}{3}[1 - e^{-4}(1 + 4)] + \frac{2}{3}[1 - e^{-5}(1 + 5)] = \frac{1}{3}(0.9084) + \frac{2}{3}(0.9596) \approx 0.9425$. Total probability over which bus, then a Poisson inside each case.

9.4 The Poisson process

ASSUMPTIONS. Events occur in time such that, for a short interval of length h,

(1) P(exactly one event in the interval) $= \lambda h + o(h)$ (proportional to the length, for a constant rate λ);

(2) P(two or more events in the interval) $= o(h)$ (no simultaneous events);

(3) the numbers of events in non-overlapping intervals are independent.

($o(h)$ means a quantity with $o(h)/h \to 0$ as $h \to 0$: negligible compared with h.)

KEY FACT. Under these assumptions, $N(t)$ = the number of events in $[0, t]$ is $\text{Poisson}(\lambda t)$. λ is the rate per unit of time; the mean number of events in a window is rate $\times$ length.

Examples from the slides: earthquakes, deaths among an insurer's policy holders, shark attacks, buses arriving at a stop. The sample path $N(t)$ is a staircase that goes up by 1 at each event.

UNITS. Convert the window to the rate's unit before multiplying. A rate of 7.2 per day over 2 hours gives $\lambda t = 7.2 \times (2/24) = 0.6$, not 14.4. A rate of 2 per hour over 10 minutes gives $2 \times (1/6) = 1/3$.

Light bulbs (slides). Bulbs burn out at rate 7.2 per day. (1) Noon to 2 p.m. is $2/24 = 1/12$ day, so $N \sim \text{Poisson}(0.6)$: $P(N > 2) = 1 - e^{-0.6}(1 + 0.6 + 0.18) \approx 0.0231$. (2) Next week: $N \sim \text{Poisson}(50.4)$, $P(N = 50) = e^{-50.4}\, 50.4^{50}/50! \approx 0.0562$. (If the calculator overflows on 50.4^{50}, compute ln of the whole thing: $-50.4 + 50 \ln 50.4 - \ln 50! \approx -2.878$, then exponentiate.)

HW8 problem 6: I-40 accidents. 3 accidents per weekday, days independent. (1) P(no accidents Monday) $= e^{-3} \approx 0.0498$. (2) Let p = P(exactly 2 on a given day) $= e^{-3}3^2/2 \approx 0.2240$. The number of days (out of 5) with exactly 2 accidents is B(5, p), so P(exactly 4 such days) $= \binom{5}{4} p^4(1 - p) \approx 5(0.2240)^4(0.7760) \approx 0.0098$. A Binomial whose p comes from a Poisson.

HW8 problems 7–8: Diane's phone. Calls arrive at rate 2 per hour. (7) In 10 minutes, $N \sim \text{Poisson}(2 \cdot 1/6) = \text{Poisson}(1/3)$: P(at least one call) $= 1 - e^{-1/3} \approx 0.2835$. (8) P(no calls in t hours) $= e^{-2t}$, which equals 0.5 when $t = (\ln 2)/2 \approx 0.3466$ hours ≈ 20.8 minutes. Breaks of at least 20.8 minutes have P(no calls) ≤ 0.5 (the homework's literal wording, "at most 0.5"); shorter breaks have P(no calls) > 0.5. Either way the number to report is the threshold $t = (\ln 2)/2$, found by taking ln of both sides.

Practice set, problem 11. Earthquakes at 4 per week; in 2 weeks $N \sim \text{Poisson}(8)$, so $P(N = 4) = e^{-8}\, 8^4/4!$: option (C). Options with e^{-4} forgot to multiply the rate by the length; option (D) divided instead of multiplying.

HOW IT IS ASKED. A rate, a time window, and a question about "exactly k", "at least one" or "more than two" events. Compute λt first, write it down, then use the Poisson pmf. A follow-up often asks "how long" or "how many" for a target probability: set up $1 - e^{-\lambda t}$ = target and solve with ln. Section 10 connects the time to the first event (Exponential) to this same process.

9.5 Check yourself

1. 100 runners, each yells with probability 0.03. Using the Poisson approximation, P(yelled at more than twice)? How many runners make P(yelled at at least once) ≥ 0.95? [HW8 P1]

Poisson(3): $1 - 8.5e^{-3} \approx 0.5768$. $n \geq \ln 20/0.03 \approx 99.9$, so $n = 100$.

2. X ~ Poisson(λ) with P(X = 1 | X ≤ 1) = 0.5. Find λ. [HW8 P2]

$\lambda/(1 + \lambda) = 0.5$, $\lambda = 1$.

3. Colds are Poisson(5) per year; a drug lowers the rate to 3 for 75% of people. A user had 2 colds; P(the drug benefits him)? [HW8 P4]

$0.75(0.2240)/[0.75(0.2240) + 0.25(0.0842)] \approx 0.8886$.

4. Clark errs on each of 400 questions with probability 0.0075, Diana with 0.003, independently. P(their total number of errors is exactly 2)? [HW8 P5]

Poisson(3) + Poisson(1.2) = Poisson(4.2): $e^{-4.2}(4.2)^2/2 \approx 0.1323$.

5. Accidents are Poisson(3) per weekday. P(none on Monday)? P(exactly 4 of the 5 weekdays have exactly 2 accidents)? [HW8 P6]

$e^{-3} \approx 0.0498$; $5p^4(1-p)$ with $p = 4.5e^{-3} \approx 0.2240$, giving ≈ 0.0098.

6. Calls arrive at 2 per hour. P(a call during a 10-minute break)? For what break length t is P(no calls) exactly 0.5, and for which t is it at most 0.5? [HW8 P7–8]

$1 - e^{-1/3} \approx 0.2835$; $e^{-2t} = 0.5$ at $t = (\ln 2)/2 \approx 0.347$ h ≈ 20.8 min, and P(no calls) ≤ 0.5 for every $t \geq 20.8$ min.

7. Earthquakes at 4 per week. P(exactly 4 in the next 2 weeks)? Brad passes 200 runners who each yell with probability 0.01; P(more than 2 yell)? [Practice 11–12]

$e^{-8}8^4/4! \approx 0.0573$; Poisson(2): $1 - 5e^{-2} \approx 0.3233$.

8. In Brad's dorm (180 others), P(at least two share his birthday), by Poisson approximation?

$\lambda = 180/365 \approx 0.4932$; $1 - e^{-\lambda}(1+\lambda) \approx 0.0881$.

9. $X \sim$ Poisson(λ) with $P(X = 2) = P(X = 3)$. Find λ.

$\lambda^2/2 = \lambda^3/6 \Rightarrow \lambda = 3$.

10. A 300-page book averages 1 typo per page. P(a given page has at least 2 typos)? Expected number of pages with no typos?

$1 - 2e^{-1} \approx 0.2642$. Each page is typo-free with probability e^{-1}, so about $300e^{-1} \approx 110$ pages.

Deck 6 10. Uniform and Exponential random variables

The idea. Two continuous "X-men". The Uniform is "a point chosen at random from an interval": every sub-interval of the same length is equally likely. The Exponential is "the waiting time until the next event" when events occur at a constant rate; it is the continuous twin of the Poisson process and it has no memory, which makes it both strange and extremely useful.

10.1 The Uniform distribution on (a, b)

Uniform(a, b).

$$f(x) = \frac{1}{b-a} \text{ for } a < x < b \text{ (0 otherwise)}; \quad F(x) = \frac{x-a}{b-a} \text{ for } a \le x \le b.$$

$$E(X) = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}, \quad P(x_1 < X < x_2) = \frac{x_2 - x_1}{b-a}.$$

Probability depends only on the length of the sub-interval, not on where it sits (shifting x_1 and x_2 by the same h changes nothing).

Mean and variance (slides). $E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$. $E(X^2) = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$. Subtracting $\frac{(a+b)^2}{4}$ and simplifying gives $\frac{(b-a)^2}{12}$. For (0, 1): mean 1/2, variance 1/12, as in section 6.

Practice set, problem 19 (chocolate bar again). The toothpick is at $p = 1/4$ and the break point U is uniform on [0, 1]. Amy's piece has length $g(U) = 1 - U$ if $U < 1/4$ (she gets [U, 1]) and $g(U) = U$ if $U > 1/4$. $P(g(U) \ge 1/2) = P(U < 1/4 \text{ and } 1 - U \ge 1/2) + P(U \ge 1/4 \text{ and } U \ge 1/2) = P(U < 1/4) + P(U \ge 1/2) = 1/4 + 1/2 = 3/4$, in (0.65, 1]. Draw the interval and mark where the piece is long; the answer is the total length of those regions.

Bus breakdown (slides, "slightly complicated"). A bus breaks down at a point X uniform on (0, 100) miles from city A. Service stations at 0, 50 and 100, or at 25, 50 and 75: which plan gives the smaller expected distance to the nearest station?

- Stations at 0, 50, 100: the distance $D_1(x)$ is x on (0, 25], 50 − x on (25, 50], x − 50 on (50, 75], 100 − x on (75, 100]. By symmetry the four pieces contribute equally: $E(D_1) = 4\int_0^{25} x \cdot \frac{1}{100}\,dx = 4 \cdot \frac{312.5}{100} = 12.5$ miles.
- Stations at 25, 50, 75: $D_2(x)$ is 25 − x on (0, 25], then triangles of half-width 12.5 around 25, 50, 75, then x − 75 on (75, 100]. $E(D_2) = \frac{1}{100}[312.5 + 4 \cdot 78.125 + 312.5] = 9.375$ miles.

The suggested plan is better. The method is the plug-in rule with a piecewise function: split the integral wherever the formula for the distance changes.

10.2 Why the Uniform matters: it generates every other distribution

Theorem (slides). Let F be a continuous, strictly increasing cdf with inverse F^{-1}.

1. If $U \sim \mathrm{Uniform}(0, 1)$, then $X = F^{-1}(U)$ has cdf F.
2. If X has cdf F, then $F(X) \sim \mathrm{Uniform}(0, 1)$.

Proof of (1): $P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$, because for a uniform $P(U \le u) = u$. Proof of (2): $P(F(X) \le z) = P(X \le F^{-1}(z)) = F(F^{-1}(z)) = z$. Punchline: a computer that can produce uniform random numbers can simulate *any* distribution: compute the cdf, draw U, output $F^{-1}(U)$. For an Exponential, $F^{-1}(u) = -\ln(1-u)/\lambda$. The

slides' "even cooler fact": the binary digits of one uniform are an infinite sequence of independent fair coin flips, so a single uniform contains infinitely many independent uniforms.

10.3 The Exponential distribution

EXPONENTIAL(λ), $\lambda > 0$.

$$f(x) = \lambda e^{-\lambda x} \text{ for } x > 0 \text{ (0 for } x < 0); \; F(x) = 1 - e^{-\lambda x}; \; P(X > x) = e^{-\lambda x}.$$

$$E(X) = \frac{1}{\lambda}, \; \mathrm{Var}(X) = \frac{1}{\lambda^2}, \; \mathrm{SD}(X) = \frac{1}{\lambda}, \; M_X(t) = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda \text{ (and } +\infty \text{ for } t \geq \lambda).$$

λ is a **rate** (events per unit time); the mean waiting time is $1/\lambda$.

Derivations (slides). $F(a) = \int_0^a \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}$ from 0 to $a = 1 - e^{-\lambda a}$. $E(X) = \int_0^\infty x\,\lambda e^{-\lambda x}\,dx = 1/\lambda$ by parts (section 1.3). $E(X^2) = \int_0^\infty x^2\,\lambda e^{-\lambda x}\,dx = 2/\lambda^2$, so $\mathrm{Var}(X) = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$. MGF: $\int_0^\infty e^{tx}\,\lambda e^{-\lambda x}\,dx = \lambda \int_0^\infty e^{-(\lambda - t)x}\,dx = \lambda/(\lambda - t)$, which needs $\lambda - t > 0$.

RATE OR MEAN?. "Exponential with mean 2 years" means $\lambda = 1/2$ per year (HW10 problem 9). "Average 43,800 hours" means $\lambda = 1/43{,}800$ (slides). "Rate $\lambda = 1$ per hour" means mean 1 hour (HW10 problem 10). Decide which you have been given before touching $e^{-\lambda x}$.

HW10 problem 9. Time to the next earthquake is Exponential with mean 2 years, so $\lambda = 1/2$. P(within one year) $= 1 - e^{-1/2} \approx 0.3935$. P(within ten years) $= 1 - e^{-5} \approx 0.9933$.

Practice set, problem 25 (the median). Laptop lifetime Exponential with mean 36 months, $\lambda = 1/36$. The median a solves $P(X \leq a) = 1 - e^{-a/36} = 1/2$, so $e^{-a/36} = 1/2$ and $a = 36 \ln 2 \approx 24.95$ months, in [20, 30). The median of an Exponential is always $(\ln 2)/\lambda \approx 0.69 \times$ the mean: the distribution is skewed right, so half the laptops die before 25 months even though the mean is 36.

Why it matters (slides). Waiting times (earthquakes, phone calls, hospital admissions); queueing theory (how businesses and internet routers behave); Markov processes (built on the memoryless property); and the Gamma distribution (sums of Exponentials) used in text mining and clustering.

10.4 The memoryless property

MEMORYLESS. If $X \sim$ Exponential(λ), then for all $s, t > 0$,

$$P(X > s + t \mid X > s) = P(X > t).$$

Having waited s already tells you nothing about how much longer you will wait. The Exponential is the only continuous distribution with this property.

Proof (slides): $P(X > s + t \mid X > s) = \dfrac{P(X > s + t \text{ and } X > s)}{P(X > s)} = \dfrac{P(X > s + t)}{P(X > s)} = \dfrac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$.
Conversely, if $G(t) = P(X > t)$ satisfies $G(s + t) = G(s)G(t)$, differentiating in s at $s = 0$ gives $G'(t) = G'(0)G(t)$, whose only solutions with $G(0) = 1$ are $G(t) = e^{-\lambda t}$ with $\lambda = -G'(0) > 0$. That is the characterization.

Eva's hard drive (slides). Lifetime Exponential with mean 43,800 hours. She has used it for 3 years and is leaving on a 4380-hour trip. P(it survives the trip | it has survived 3 years) $= P(X > 4380) = e^{-4380/43{,}800} = e^{-0.1} \approx 0.9048$, by memorylessness; the three years of use are irrelevant. If the lifetime were not Exponential (a drive that wears out), the age would matter and you would have to compute $P(X > s + t)/P(X > s)$ from the actual cdf.

HW10 problem 10. Repair time Exponential with rate 1 per hour. P(takes at least 10 hours | has exceeded 9 hours) = $P(X \geq 1) = e^{-1} \approx 0.3679$. (Computing (e^{-10}/e^{-9}) directly gives the same thing.)

HOW IT IS ASKED. "Given that it has lasted s, probability it lasts at least t more": the answer is $e^{-\lambda t}$, full stop. If the question instead asks for "at least a total of s + t" you still get $e^{-\lambda t}$; the only way to go wrong is to forget to subtract s.

10.5 Connection to the Poisson process

TIME TO THE FIRST EVENT. If events form a Poisson process with rate λ and X is the time of the first event, then $\{X > t\} = \{N(t) = 0\}$, so $P(X > t) = P(N(t) = 0) = e^{-\lambda t}$: $X \sim \text{Exponential}(\lambda)$. The same holds for the time between any two consecutive events, and these gaps are independent.

So HW8 problem 8 ($P(\text{no calls in t hours}) = e^{-2t}$) and "the wait for the next call is Exponential(2)" are the same statement. With calls at 2 per hour, the expected wait is 1/2 hour and $P(\text{wait} > 30 \text{ min}) = e^{-1} \approx 0.3679$.

10.6 Check yourself

1. $X \sim$ Uniform(0, 30). Find E(X), Var(X), and $P(X \geq 25 \mid X > 20)$. [HW10 P4]

15; 900/12 = 75; (5/30)/(10/30) = 1/2.

2. The toothpick is at 1/4 and the break point U is uniform on [0, 1]. P(Amy's piece has length ≥ 1/2)? [PRACTICE 19]

$P(U < 1/4) + P(U \geq 1/2) = 3/4$.

3. With service stations only at the two ends (0 and 100), what is the expected distance from a uniform breakdown point to the nearest station?

$D = \min(x, 100 - x)$: $E(D) = 2 \int_0^{50} x/100 \, dx = 2(12.5) = 25$ miles.

4. Time to the next earthquake is Exponential with mean 2 years. P(within 1 year)? P(within 10 years)? [HW10 P9]

$1 - e^{-0.5} \approx 0.3935$; $1 - e^{-5} \approx 0.9933$.

5. Repair time is Exponential with rate 1 per hour. P(at least 10 hours | more than 9 hours)? [HW10 P10]

$e^{-1} \approx 0.3679$.

6. Laptop lifetime Exponential with mean 36 months. Median? P(lasts longer than the mean)? [PRACTICE 25]

$36 \ln 2 \approx 24.95$ months. $P(X > 36) = e^{-1} \approx 0.3679$ (always, for any Exponential).

7. Hard drive lifetime Exponential with mean 43,800 hours; after 3 years of use, P(it survives a further 4380 hours)?

$e^{-0.1} \approx 0.9048$, regardless of the 3 years.

8. Waiting time Exponential with mean 10 minutes. P(the wait is between 5 and 15 minutes)?

$e^{-0.5} - e^{-1.5} \approx 0.6065 - 0.2231 = 0.3834$.

9. U ~ Uniform(0, 1). What is the distribution of $X = -\ln(U)/\lambda$?

$P(X > t) = P(U < e^{-\lambda t}) = e^{-\lambda t}$: Exponential($\lambda$). ($1 - U$ is also uniform, so $-\ln(1 - U)/\lambda$ works too.)

10. Buses arrive as a Poisson process at 3 per hour. Expected wait for the next bus? P(wait more than 20 minutes)?

Wait ~ Exponential(3 per hour): mean 1/3 hour = 20 minutes; $P(\text{wait} > 1/3 \text{ h}) = e^{-1} \approx 0.3679$.

11. A Uniform(a, b) variable has mean 5 and variance 3. Find a and b.

$(b - a)^2/12 = 3$ gives $b - a = 6$; $a + b = 10$; so $a = 2$, $b = 8$.

Deck 7 11. Geometric distribution, joint distributions and linearity of expectation

The idea. Deck 7 does three things. It adds one more "famous" discrete distribution, the Geometric (trials until the first success). It starts what the slides call the most important aspect of the course: relationships between random variables, described by a joint distribution. And it proves the "super duper useful fact" that $E(X + Y) = E(X) + E(Y)$ with no conditions at all, which, combined with indicator variables, computes expectations of very complicated counts in two lines (coat matching, coupon collecting, Homework 9).

11.1 The Geometric distribution

Geometric(p). X is geometric with parameter $p \in (0, 1)$ if it takes values 1, 2, 3, … with

$$P(X = n) = (1 - p)^{n-1} p, \; n = 1, 2, \ldots$$

Interpretation (slides): the number of trials until the first success in independent Bernoulli(p) trials. Then

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t} \text{ for } t < -\ln(1 - p), \; E(X) = \frac{1}{p}, \; \mathrm{Var}(X) = \frac{1 - p}{p^2}, \; P(X > k) = (1 - p)^k.$$

Where the formulas come from (slides). "First success on trial n" is $n - 1$ failures then a success, and the trials are independent. The probabilities add to 1 by the geometric series: $p \, \Sigma_{n \geq 1}(1 - p)^{n-1} = p/(1 - (1 - p)) = 1$. "X > k" means the first k trials all failed. For the moments the slides use the MGF:

$$M_X(t) = \Sigma_{k \geq 1} \, e^{tk}(1 - p)^{k-1} p = pe^t \, \Sigma_{k \geq 1} \, [e^t(1 - p)]^{k-1} = \frac{pe^t}{1 - (1 - p)e^t},$$

which is finite when $(1 - p)e^t < 1$, i.e. $t < -\ln(1 - p)$. Differentiating, $M'(t) = \frac{pe^t}{(1 - (1 - p)e^t)^2}$, so $E(X) = M'(0) = p/p^2 = 1/p$. One more derivative gives $E(X^2) = M''(0) = (2 - p)/p^2$, hence $\mathrm{Var}(X) = (2 - p)/p^2 - 1/p^2 = (1 - p)/p^2$. (As for the Binomial and Poisson, direct summation gives the same results.) Intuition for the mean: a success happens one time in $1/p$ trials, so you wait $1/p$ trials on average.

Two conventions. The course (and the homework) counts the trial of the first success: X = 3 for TTH, values 1, 2, 3, …, mean $1/p$. The textbook's Geom(p) counts *failures before* the first success (values 0, 1, 2, …, mean $(1 - p)/p$). Read the definition in the problem; the course's problems use the "number of trials" version, as above.

Memoryless. $P(X > m + k \mid X > m) = (1 - p)^{m+k}/(1 - p)^m = (1 - p)^k = P(X > k)$: having failed m times tells you nothing about the future, just like the Exponential.

Practice set, problem 6. Each kidney matches with probability 0.1 independently; X = number tested until a match. $E(X) = 1/0.1 = 10$, in [6, 12).

HW6 problem 8(1), revisited. P(Eva's first mistake is on order 11) $= (1/2)^{10}(1/2)$: the Geometric(1/2) pmf at n = 11.

HW9 problem 7: who gets heads first. Your coin has $P(H) = 3/4$, your friend's has 1/4; X and Y are the numbers of flips each needs, independent. $P(X < Y)$? Condition on the value of X:

$$P(X < Y) = \Sigma_{k \geq 1} \, P(X = k) \, P(Y > k) = \Sigma_{k \geq 1} \, (1/4)^{k-1}(3/4) \cdot (3/4)^k = \frac{9}{16} \, \Sigma_{k \geq 1} \, (3/16)^{k-1} =$$

$$\frac{9/16}{1 - 3/16} = \frac{9}{13} \approx 0.6923.$$

First-step check: in a round where both flip, you win outright if you get H and your friend gets T (probability (3/4)(3/4) = 9/16); the round is a do-over if both get T ((1/4)(3/4) = 3/16); otherwise your friend has H and you have not beaten them. So p = 9/16 + (3/16)p, p = 9/13. Both methods are always available; use whichever you set up faster.

11.2 Joint distributions: cdf, pmf, marginals

The slides motivate joint distributions with pairs like (Netflix price, Amazon price), (cholesterol, blood pressure), (rainfall in NC, rainfall in VA): most of statistics (regression, time series, factor models) and most of probability (Markov chains, stochastic processes) is about collections of random variables and how they relate.

JOINT CDF. For two random variables, $F(a, b) = P(X \le a, Y \le b)$, $-\infty < a, b < \infty$. Every probability question about X and Y can be answered from F. It contains the individual cdfs: $F_X(a) = \lim_{b\to\infty} F(a, b)$. And, for example (slides),

$$P(X > a, Y > b) = 1 - P(X \le a \text{ or } Y \le b) = 1 - [F_X(a) + F_Y(b) - F(a, b)] = 1 - F_X(a) - F_Y(b) + F(a, b).$$

JOINT PMF AND MARGINALS (DISCRETE CASE).

$$p_{X,Y}(x, y) = P(X = x, Y = y),\ p_X(x) = P(X = x, Y \text{ takes any value}) = \Sigma_y\, p_{X,Y}(x, y),\ p_Y(y) = \Sigma_x\, p_{X,Y}(x, y).$$

The marginal pmf of X is just the pmf of X; the word "marginal" comes from writing the row and column totals in the margins of the table. For more than two variables, sum over every coordinate except the one you want to keep. (The other case, jointly continuous X and Y with a joint density, is deck 9 and is not on this midterm.)

Extra context beyond the slides: the conditional pmf $p_{X|Y}(x \mid y) = p_{X,Y}(x, y)/p_Y(y)$ is the multiplication rule in pmf form, so $p_{X,Y}(x, y) = p_Y(y)\, p_{X|Y}(x \mid y)$. Homework 9 problem 2 and practice problem 22 describe one variable "given" the other and need exactly this.

Problem 1 from the slides: two dice, X = the larger value, Y = the sum. All 36 ordered pairs (d_1, d_2) are equally likely. For the maximum to be x and the sum to be y, one die must show x and the other $y - x \le x$. If $y - x < x$ there are two ordered pairs, $(x, y - x)$ and $(y - x, x)$; if $y = 2x$ there is one, (x, x). So

$$p(x, y) = \frac{2}{36} \text{ for } x + 1 \le y \le 2x - 1,\ p(x, 2x) = \frac{1}{36}, \text{ and 0 otherwise.}$$

p(x, y)	y=2	3	4	5	6	7	8	9	10	11	12	$p_X(x)$
x = 1	1											1/36
x = 2		2	1									3/36
x = 3			2	2	1							5/36
x = 4				2	2	2	1					7/36
x = 5					2	2	2	2	1			9/36
x = 6						2	2	2	2	2	1	11/36
$p_Y(y)$	1	2	3	4	5	6	5	4	3	2	1	36/36

(Entries are in 36ths; blank = 0.) The row totals are the pmf of the maximum, $(2x-1)/36$, and the column totals are the familiar pmf of the sum. Reading probabilities off the table: $P(X = 4, Y = 7) = 2/36$, $P(Y = 7 \mid X = 4) = (2/36)/(7/36) = 2/7$, $P(X \le 2, Y \le 3) = F(2, 3) = 3/36$.

A smaller table. Toss a fair coin three times; X = heads in the first two tosses, Y = total heads:

$p(x, y)$	$y = 0$	$y = 1$	$y = 2$	$y = 3$	$p_X(x)$
$x = 0$	1/8	1/8	0	0	2/8
$x = 1$	0	2/8	2/8	0	4/8
$x = 2$	0	0	1/8	1/8	2/8
$p_Y(y)$	1/8	3/8	3/8	1/8	1

$P(X = Y) = 1/8 + 2/8 + 1/8 = 1/2$ (the third toss is tails); $P(X = 1 \mid Y = 2) = (2/8)/(3/8) = 2/3$. This table returns in section 12 as an example of dependent variables.

Practice set, problem 14 (sampling without replacement). 3 sophomores, 4 juniors, 5 seniors; 3 are chosen at random; X = sophomores chosen, Y = juniors chosen. Every triple is equally likely, so

$$P(X = x, Y = y) = \frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}}, \text{ and } P(X = 1, Y = 1) = \frac{3 \cdot 4 \cdot 5}{\binom{12}{3}} = \frac{60}{220} \approx 0.2727\text{: option (B).}$$

Option (D), (3/12)(4/12), treats "one sophomore" and "one junior" as two independent single draws, which is wrong twice over: the draws are without replacement, and even with replacement the probability of one of each class in three draws would be the multinomial $3! \cdot (3/12)(4/12)(5/12) \approx 0.2083$, not a bare product of two fractions.

HW9 problem 2: two doctors. H = 1 if Dr. Hibbert performs the n surgeries (probability p), else H = 0. Given H = 1, each surgery succeeds with probability a independently, so $X \mid H = 1 \sim \text{Binomial}(n, a)$; given H = 0, $X \mid H = 0 \sim \text{Binomial}(n, b)$. By the multiplication rule the joint pmf is

$$P(H = 1, X = k) = p\binom{n}{k} a^k(1-a)^{n-k},\ P(H = 0, X = k) = (1-p)\binom{n}{k} b^k(1-b)^{n-k},\ k = 0, \ldots, n.$$

The marginal of X is the sum of the two rows: a mixture of two Binomials. For E(X), condition on the doctor: $E(X) = p \cdot na + (1-p) \cdot nb = n[pa + (1-p)b]$. For $E(X^2)$, use $E(Y^2) = \text{Var}(Y) + (E\,Y)^2$ for each Binomial: $E(X^2) = p[na(1-a) + n^2a^2] + (1-p)[nb(1-b) + n^2b^2]$. Then $\text{Var}(X) = E(X^2) - (E\,X)^2$, which simplifies to $n[pa(1-a) + (1-p)b(1-b)] + n^2p(1-p)(a-b)^2$: the average within-doctor variance plus the extra variance from not knowing which doctor.

Practice set, problem 22 (a product of dependent variables). Shankar gets X questions with $P(X = 0, 1, 2) = 0.2, 0.4, 0.4$; each question is answered inadequately with probability 1/5 independently, so $Y \mid X = x \sim \text{Binomial}(x, 1/5)$. $P(XY \le 1)$: go case by case on X.

- $X = 0$: $XY = 0 \le 1$ always. Contributes 0.2.
- $X = 1$: $Y \in \{0, 1\}$ so $XY = Y \le 1$ always. Contributes 0.4.
- $X = 2$: $XY = 2Y \le 1$ only if $Y = 0$, probability $(4/5)^2 = 0.64$. Contributes $0.4 \times 0.64 = 0.256$.

Total 0.856, in (0.6, 1.0]. When one variable's distribution is described "given" the other, the joint pmf is $p_X(x)\ p_{Y|X}(y \mid x)$ and you sum over the cases.

11.3 Expectations from a joint pmf, and linearity

PLUG-IN RULE, TWO VARIABLES. If X, Y have joint pmf $p_{X,Y}$ and g(x, y) is any function,

$$E[g(X, Y)] = \Sigma_{x,y}\, g(x, y)\, p_{X,Y}(x, y).$$

With g(x, y) = x this gives $\Sigma_x\, x\, p_X(x) = E(X)$: computing E(X) from the joint pmf or from the marginal gives the same answer.

THE SUPER DUPER USEFUL FACT. For any random variables, independent or not,

$$E(X + Y) = E(X) + E(Y), \text{ and } E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

Proof (slides, equation (1)): $\Sigma_{x,y}(x + y)\, p(x, y) = \Sigma_x\, x\, \Sigma_y\, p(x, y) + \Sigma_y\, y\, \Sigma_x\, p(x, y) = \Sigma_x\, x\, p_X(x) + \Sigma_y\, y\, p_Y(y) = E(X) + E(Y)$. Only the marginals enter, so dependence between X and Y is irrelevant; the n-variable version follows by induction from the two-variable case (in class: $Y = X_2 + \cdots + X_n$). Variances do *not* add this freely (section 12).

Immediate payoff: a Binomial(n, p) is a sum of n Bernoulli(p) variables, so its mean is np with no computation. In the coin table above, E(Y) = E(X) + E(third toss) = 1 + 1/2 = 3/2.

11.4 The power of indicator random variables

COUNTING WITH INDICATORS (SLIDES). If X is the number of the events $A_1, \ldots, A_n$ that occur, then $X = I_{A_1} + \cdots + I_{A_n}$, where $I_{A_i} = 1$ if A_i occurs and 0 if not, so

$$E(X) = \Sigma_{i=1}^{n} E(I_{A_i}) = \Sigma_{i=1}^{n} [1 \cdot P(A_i) + 0 \cdot P(A_i^c)] = \Sigma_{i=1}^{n} P(A_i),$$

whether or not the events are independent. You never need the distribution of X.

Ten students (slides). Student i does well with probability p_i. With $X_i = 1$ if student i does well, the number doing well is $X = \Sigma\, X_i$ and $E(X) = \Sigma_{i=1}^{10} p_i$. No independence was assumed.

Coat matching (slides). Thirty guests' coats are handed back at random, so each guest gets their own coat with probability 1/30. M = number of guests who get their own coat $= \Sigma_{i=1}^{30} I_i$, and $E(M) = 30 \cdot (1/30) = 1$. The indicators are dependent (if 29 guests have the right coat, so does the 30th), and the answer is exactly 1 for any number of guests.

The coupon collector problem (slides). Each cereal box contains one of 100 X-men cards, chosen uniformly and independently. X = number of boxes needed to collect all 100. Break the wait into pieces: T_i = the number of boxes bought while you own exactly i distinct cards, until a new one appears, i = 0, 1, ..., 99. Each box is "new" with probability (100 − i)/100, so T_i is Geometric with p = (100 − i)/100 and $E(T_i) = 100/(100 - i)$. Then $X = T_0 + T_1 + \cdots + T_{99}$ and

$$E(X) = \Sigma_{i=0}^{99} \frac{100}{100 - i} = 100\, (1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{100}) = 100\, H_{100} \approx 518.74 \text{ boxes.}$$

At \$2.50 a box, Z = 2.5X and $E(Z) = 2.5 \cdot 518.74 \approx \$1{,}296.85$. The slides' punchline: "E(sum of a bunch of random variables) = sum of the individual expectations" was the whole tool; the problem looks made up but shows up in computer science and genetics (see the course "Extras").

HW9 problem 1: five students, four colleges. Each of 5 students independently picks UNC, Duke, NC State or Hogwarts-NC with probabilities 0.6, 0.1, 0.1, 0.2. I_{UNC} is the indicator that exactly 2 of the 5 choose UNC, and so on; X = number of schools that get exactly 2 students.

(1) The number choosing a given school is Binomial(5, p_{school}), so $E(I_{UNC}) = P(\text{exactly } 2) = \binom{5}{2}(0.6)^2(0.4)^3 = 0.2304$; $E(I_{Duke}) = \binom{5}{2}(0.1)^2(0.9)^3 = 0.0729$; $E(I_{NCS}) = 0.0729$; $E(I_{HNC}) = \binom{5}{2}(0.2)^2(0.8)^3 = 0.2048$.

(2) $X = I_{UNC} + I_{Duke} + I_{NCS} + I_{HNC}$.

(3) $E(X) = 0.2304 + 0.0729 + 0.0729 + 0.2048 = 0.581$.

HW9 problem 3: Halloween houses. 12 houses in a row, 7 give candy, every arrangement equally likely; X = number of adjacent pairs with different behavior. Let $I_i = 1$ if houses i and i + 1 differ, i = 1, …, 11. For any pair, P(different) = P(C then N) + P(N then C) = $\frac{7}{12} \cdot \frac{5}{11} + \frac{5}{12} \cdot \frac{7}{11} = \frac{70}{132} = \frac{35}{66}$. So $E(X) = 11 \cdot \frac{35}{66} = \frac{35}{6} \approx 5.83$. The pmf of X would be a nightmare; the expectation takes two lines.

Practice set, problem 20: sad schools. Four athletes choose independently among UNC, Duke, NC State with the given probabilities. Sad = number of schools that get nobody. $E(\text{Sad}) = \Sigma_{\text{schools}}\, P(\text{nobody picks it})$: UNC (0.5)(0.2)(0.7)(0.1) = 0.007; Duke (0.7)(0.9)(0.6)(0.95) = 0.3591; NC State (0.8)(0.9)(0.7)(0.95) = 0.4788. E(Sad) = 0.8449, in (0.6, 1.2]. Each probability is a product because the athletes decide independently; the three indicators are not independent, and again that does not matter.

How it is asked. "Let X be the number of … . Find E(X)." If X counts events, write X as a sum of indicators, compute one probability per event, add. The homework comment calls this "the famous method of indicators" for a reason. If X is a waiting time that can be split into stages (coupon collector), write it as a sum of geometric waits.

11.5 The multinomial distribution

Multinomial (slides). n independent trials; each trial has exactly r possible outcomes, with probabilities p_1, …, p_r ($\Sigma\, p_i = 1$) that do not change from trial to trial. X_i = number of trials with outcome i, so $X_1 + \cdots + X_r = n$. For integers $n_i \geq 0$ with $n_1 + \cdots + n_r = n$,

$$P(X_1 = n_1, \ldots, X_r = n_r) = \frac{n!}{n_1!\, n_2! \cdots n_r!}\, {p_1}^{n_1}\, {p_2}^{n_2} \cdots {p_r}^{n_r},$$

and 0 for any other count vector.

Why (as written in class): a specific sequence of outcomes with n_i of type i has probability $\Pi\, {p_i}^{n_i}$, and the multinomial coefficient counts such sequences (the "labeled groups" count from Midterm 1). With r = 2 it is the Binomial. Each X_i on its own is Binomial(n, p_i) (a trial is "type i" or not), and the count of a union of categories, $X_i + X_j$, is Binomial(n, $p_i + p_j$).

HW9 problem 4: three candidates. 35% Sauron, 45% Voldemort, 20% Saruman; sample 4 voters.

(1) $P(X_A = 0, X_V = 1, X_S = 3) = \frac{4!}{0!\, 1!\, 3!}\, (0.35)^0(0.45)^1(0.2)^3 = 4(0.45)(0.008) = 0.0144$.

(2) $Y = X_A + X_S$ = number of non-Voldemort voters ~ Binomial(4, 0.55): E(Y) = 2.20, Var(Y) = 4(0.55)(0.45) = 0.99.

Practice set, problem 23: ramen. Five customers; $P(\text{2 miso, 2 tonkotsu, 1 shoyu}) = \frac{5!}{2!\, 2!\, 1!}\, (0.5)^2(0.3)^2(0.1) = 30(0.25)(0.09)(0.1) = 0.0675$, in (0.0025, 0.1]. (The unused category, shio, simply has $n_i = 0$.)

11.6 Check yourself

1. Each kidney matches with probability 0.1; X = number tested until the first match. E(X)? P(X > 5)? P(X = 7 | X > 5)? [Practice 6]

10; $0.9^5 \approx 0.5905$; by memorylessness $P(X = 2) = 0.9(0.1) = 0.09$.

2. Your coin: P(H) = 3/4; friend's: 1/4. Each flips until the first H. P(you need strictly fewer flips)? [HW9 P7]

$9/13 \approx 0.6923$.

3. Two dice, X = larger value, Y = sum. Find P(X = 5, Y = 8), P(X = 3), and P(Y = 7 | X = 4).

2/36 (pairs (5, 3) and (3, 5)); 5/36; (2/36)/(7/36) = 2/7.

4. Write $M_X(t)$ for a Geometric(p) and use it to get E(X). For which t is it finite?

$pe^t/(1 - (1 - p)e^t)$; $M'(0) = p/p^2 = 1/p$; $t < -\ln(1 - p)$.

5. 3 sophomores, 4 juniors, 5 seniors; choose 3 at random; X = sophomores, Y = juniors. P(X = 1, Y = 1)? P(X = 0, Y = 0)? [Practice 14]

$60/220 \approx 0.2727$; $\binom{5}{3}/220 = 10/220 \approx 0.0455$.

6. Toss a fair coin 3 times; X = heads in the first two, Y = total heads. P(Y = 2 | X = 1)? E(XY)?

$(2/8)/(4/8) = 1/2$. $E(XY) = 1 \cdot 1 \cdot (2/8) + 1 \cdot 2 \cdot (2/8) + 2 \cdot 2 \cdot (1/8) + 2 \cdot 3 \cdot (1/8) = 16/8 = 2$.

7. Dr. Hibbert (probability p) succeeds with probability a per surgery, Dr. Nick with b; n surgeries by one doctor. Write P(H = 1, X = k) and find E(X). [HW9 P2]

$p\binom{n}{k}a^k(1 - a)^{n-k}$; $E(X) = n[pa + (1 - p)b]$.

8. X questions with P(0, 1, 2) = 0.2, 0.4, 0.4; each answered badly with probability 1/5; Y = number answered badly. P(XY ≤ 1)? [Practice 22]

$0.2 + 0.4 + 0.4(0.64) = 0.856$.

9. Five students each pick UNC/Duke/NCS/HNC with probabilities 0.6/0.1/0.1/0.2. E(number of schools receiving exactly 2 students)? [HW9 P1]

$0.2304 + 0.0729 + 0.0729 + 0.2048 = 0.581$.

10. 12 houses, 7 give candy, random arrangement. E(number of adjacent pairs that differ)? [HW9 P3]

$11 \cdot 35/66 = 35/6 \approx 5.83$.

11. Four athletes choose UNC/Duke/NCS with the probabilities in practice problem 20. E(number of schools with no athlete)? [Practice 20]

$0.007 + 0.3591 + 0.4788 = 0.8449$.

12. A cereal brand uses 6 different cards, equally likely and independent across boxes. Expected number of boxes to collect all 6? And 30 guests' coats handed back at random: expected number of correct coats?

$6(1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6) = 6(2.45) = 14.7$ boxes. Exactly 1 coat.

13. 35% Sauron, 45% Voldemort, 20% Saruman; sample 4. P(0 Sauron, 1 Voldemort, 3 Saruman)? E and Var of the number of non-Voldemort voters? [HW9 P4]

0.0144; Binomial(4, 0.55): 2.2 and 0.99.

14. Five ramen customers with preferences 0.1 shoyu, 0.1 shio, 0.5 miso, 0.3 tonkotsu. P(2 miso, 2 tonkotsu, 1 shoyu)? [Practice 23]

$30(0.25)(0.09)(0.1) = 0.0675$.

Deck 8 12. Independent random variables, covariance and correlation

The idea. Deck 8 first isolates the special case where one random variable carries no information about another: independence. It makes products behave ($E[g(X)h(Y)] = E[g(X)]E[h(Y)]$), makes sums behave (MGFs multiply, so Binomials and Poissons add), and makes variances add. The deck then returns to general, possibly dependent, variables and measures their *linear* relationship with covariance and correlation, which gives "super powerful formulas and rules for computing variances of sums".

12.1 Independence of random variables

Definition (slides). X and Y are independent if for any sets A and B

$$P(X \in A,\, Y \in B) = P(X \in A)\, P(Y \in B),$$

that is, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent. Otherwise X and Y are dependent. Equivalent condition 1: $F(a, b) = F_X(a)F_Y(b)$ for all a, b. Equivalent condition 2 (discrete case):

$$p(x, y) = p_X(x)\, p_Y(y) \text{ for all } x, y,$$

and this is also equivalent to $p(x, y) = h(x)g(y)$ for some functions h and g (the pmf factors). For n variables: $p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdot \cdot \cdot p_{X_n}(x_n)$.

Why condition 2 is enough (in class): $P(X \in A, Y \in B) = \Sigma_{x\in A}\, \Sigma_{y\in B}\, p(x, y) = \Sigma_{x\in A}\, \Sigma_{y\in B}\, p_X(x)p_Y(y) = [\Sigma_{x\in A}\, p_X(x)]\, [\Sigma_{y\in B}\, p_Y(y)] = P(X \in A)P(Y \in B)$. And X, Y are *not* independent as soon as there are sets A, B with $P(X \in B \mid Y \in A) \neq P(X \in B)$.

Two quick tests. (1) In a table: independent exactly when every cell equals its row total times its column total; one failing cell is enough to prove dependence. (2) If the pmf is a product h(x)g(y) on a *rectangular* set of values, they are independent; if the set of possible y depends on x (a triangle, not a rectangle), they are dependent whatever the formula.

Final Exam 2011 problem (slides). Two balls are drawn without replacement from an urn with 5 white and 8 red balls. $X_1 = 1$ if the first ball is red (else 0), $X_2 = 1$ if the second is red. Joint pmf by the multiplication rule:

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$	marginal of X_1
$x_1 = 0$	(5/13)(4/12) = 5/39	(5/13)(8/12) = 10/39	5/13
$x_1 = 1$	(8/13)(5/12) = 10/39	(8/13)(7/12) = 14/39	8/13
marginal of X_2	5/13	8/13	1

Both marginals are the same (the second ball is red with probability 8/13 too), but $P(X_1 = 1)P(X_2 = 1) = (8/13)^2 = 64/169 \approx 0.379$ while $P(X_1 = 1, X_2 = 1) = 14/39 \approx 0.359$. One cell fails, so X_1 and X_2 are dependent, as sampling without replacement should be.

Pet boarding: thinning a Poisson (slides). $N \sim$ Poisson(λ) dogs arrive in a day; each independently is "crazy" with probability p. Cr = number of crazy puppies, No = number of normal dogs, Cr + No = N. For $c, d \geq 0$, condition on $N = c + d$; given N, Cr is Binomial(N, p):

$$P(Cr = c,\ No = d) = P(N = c + d)\ P(Cr = c \mid N = c + d) = e^{-\lambda}\ \frac{\lambda^{c+d}}{(c + d)!}\ \cdot\ \frac{(c + d)!}{c!\ d!}\ p^c(1 - p)^d$$
$$= [e^{-\lambda p}\ \frac{(\lambda p)^c}{c!}]\ \cdot\ [e^{-\lambda(1-p)}\ \frac{(\lambda(1 - p))^d}{d!}],$$

using $e^{-\lambda} = e^{-\lambda p}e^{-\lambda(1-p)}$. The joint pmf is a product of a Poisson(λp) pmf in c and a Poisson($\lambda(1 - p)$) pmf in d, so Cr and No are independent Poissons with means λp and $\lambda(1 - p)$. This is the "factorization test" in action, and it is surprising: the two counts share the same random N and are still independent.

Practice set, problem 13. $p(x, y) = (x^2 + y^2)/140$ for x = 1, 2, 3 and y = −x, …, x. Dependent, for two separate reasons: the range of y depends on x, and $x^2 + y^2$ is a sum, not a product h(x)g(y). Option (B).

The coin table of section 11.2. X = heads in tosses 1–2, Y = total heads: $p(0, 3) = 0$ but $p_X(0)p_Y(3) = (2/8)(1/8) \neq 0$. Dependent ($Y \geq X$ always).

INGREDIENT 2 (SLIDES): PRODUCTS OF INDEPENDENT VARIABLES. If X and Y are independent and g, h are functions,

$$E[g(X)h(Y)] = E[g(X)]\ E[h(Y)],$$

and the same for any number of independent variables. Proof: $\Sigma_{x,y}\ g(x)h(y)\ p_X(x)p_Y(y) = [\Sigma_x\ g(x)p_X(x)][\Sigma_y\ h(y)p_Y(y)]$. In particular E(XY) = E(X)E(Y), and functions of independent variables are independent.

HW9 problem 8: Bert and Ernie's sprints. X ~ Binomial(5, 2/3) (from its MGF), Y ~ Poisson(2), independent; Z = XY. E(Z) = E(X)E(Y) = (10/3)(2) = 20/3 ≈ 6.67. For the variance, $Var(Z) = E(Z^2) - (E\ Z)^2$ with $E(Z^2) = E(X^2)E(Y^2)$ (independence again, applied to X^2 and Y^2). $E(X^2) = Var(X) + (E\ X)^2 = 10/9 + 100/9 = 110/9$; $E(Y^2) = 2 + 4 = 6$. So $E(Z^2) = (110/9)(6) = 220/3$ and $Var(Z) = 220/3 - 400/9 = 260/9 \approx 28.89$.

HW10 problem 11. X ~ N(0, 1) and Y ~ Exponential(1) independent: $E[|X|(Y + 1)^2] = E|X|\ \cdot\ E[(Y + 1)^2]$. $E|X| = \sqrt{(2/\pi)} \approx 0.7979$ (section 13.3). $E[(Y + 1)^2] = E(Y^2) + 2E(Y) + 1 = (1 + 1) + 2 + 1 = 5$, using $E(Y^2) = Var(Y) + (E\ Y)^2 = 1 + 1$. Answer: $5\sqrt{(2/\pi)} \approx 3.989$.

12.2 Sums of independent random variables

Why sums (slides): a sample of 100 Chapel Hill voters with $X_i = 1$ if voter i supports the candidate gives the total $\Sigma\ X_i \sim$ Binomial(100, p); a study comparing two universities bases its conclusion on the sample averages $(1/40)\ \Sigma\ X_i$ and $(1/30)\ \Sigma\ Y_j$. Two methods find the distribution of a sum.

METHOD 1: DIRECT (CONVOLUTION). For independent discrete X, Y: $P(X + Y = n) = \Sigma_k\ P(X = k)\ P(Y = n - k)$. Example (slides): $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$:

$$P(X + Y = n) = \Sigma_{k=0}{}^{n}\ e^{-\lambda_1}\frac{\lambda_1{}^k}{k!}\ e^{-\lambda_2}\frac{\lambda_2{}^{n-k}}{(n - k)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\ \Sigma_{k=0}{}^{n}\ \binom{n}{k}\ \lambda_1{}^k\lambda_2{}^{n-k} = e^{-(\lambda_1+\lambda_2)}\ \frac{(\lambda_1 + \lambda_2)^n}{n!},$$

by the binomial theorem: $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Method 2: moment generating functions. Ingredient 1: an MGF finite on an open interval around 0 determines the distribution. Ingredient 2: $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for independent variables. Combining them, for independent X_1, …, X_n and $S = X_1 + \cdots + X_n$,

$$M_S(t) = E(e^{tX_1}\, e^{tX_2} \cdots e^{tX_n}) = M_{X_1}(t)\, M_{X_2}(t) \cdots M_{X_n}(t).$$

Recipe: compute the individual MGFs, multiply, and see whether the product is the MGF of a known distribution.

- **Binomials with the same p (warning-sign slide).** $X_i \sim \text{Bin}(n_i, p)$ independent: $M_S(t) = \Pi\,(1 - p + pe^t)^{n_i} = (1 - p + pe^t)^{n_1 + \cdots + n_m}$, so $S \sim \text{Bin}(n_1 + \cdots + n_m, p)$. (Different p's do not combine into a Binomial.)
- **Poissons (warning-sign slide).** $X_i \sim \text{Pois}(\lambda_i)$ independent: $M_S(t) = e^{(\lambda_1 + \cdots + \lambda_m)(e^t - 1)}$, so $S \sim \text{Pois}(\lambda_1 + \cdots + \lambda_m)$, agreeing with Method 1.
- **Normals** (section 13.2): $N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
- Sums of Geometrics are not Geometric and sums of Exponentials are not Exponential (their MGF products are not of the same form).

HW8 problem 5 was this: Poisson(3) + Poisson(1.2) = Poisson(4.2), so $P(\text{sum} = 2) = e^{-4.2}(4.2)^2/2 \approx 0.1323$. By Method 1 instead: P(0)P(2) + P(1)P(1) + P(2)P(0) with the two pmfs gives the same number, more slowly.

12.3 Maximum and minimum of independent variables

Extra context beyond the slides, needed for HW9 problem 6 and practice problem 17.

Use the cdf. For independent X_1, …, X_n:

$$P(\max \le a) = P(\text{all } X_i \le a) = \Pi_i\, P(X_i \le a),\ P(\min > a) = P(\text{all } X_i > a) = \Pi_i\, P(X_i > a).$$

Then $P(\max = k) = P(\max \le k) - P(\max \le k - 1)$ for integer values.

HW9 problem 6. Twenty rolls of a fair four-sided die; Y = the maximum. Values 1, 2, 3, 4. $P(Y \le i) = (i/4)^{20}$, so $P(Y = i) = (i/4)^{20} - ((i-1)/4)^{20}$: about 9×10^{-13}, 9.5×10^{-7}, 0.00317, 0.99683. With 20 rolls the maximum is almost surely 4.

Practice set, problem 17. $P(\max(X, Y) \le 1) = P(X \le 1)P(Y \le 1) = 2e^{-1} \cdot 3e^{-2} = 6e^{-3} \approx 0.2987$ for independent Poisson(1) and Poisson(2).

Minimum of Exponentials. $X \sim \text{Exp}(\lambda_1)$, $Y \sim \text{Exp}(\lambda_2)$ independent: $P(\min > t) = e^{-\lambda_1 t}e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t}$, so the minimum is Exponential($\lambda_1 + \lambda_2$).

12.4 Covariance

Covariance (warning-sign slide).

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - E(X)E(Y).$$

Expanding the product: $E[XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y] = E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y$.

Meaning (slides). Covariance signals the tendency of two variables to go up or down together relative to their means: positive (height and weight of a random person), negative (X up, Y tends down). Everything is random, so one can only talk about tendency. (For spurious correlations, the slides point to tylervigen.com.)

PROPERTIES.

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$; $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$; $\mathrm{Cov}(X, c) = 0$ for a constant.
- $\mathrm{Cov}(aX + b, Y) = a\,\mathrm{Cov}(X, Y)$, and more generally $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$. (In class: $E[(aX + b)Y] - E(aX + b)E(Y) = aE(XY) + bE(Y) - aE(X)E(Y) - bE(Y)$.)
- Bilinear: $\mathrm{Cov}(\Sigma_i X_i, \Sigma_j Y_j) = \Sigma_i \Sigma_j \mathrm{Cov}(X_i, Y_j)$. Expand like a product of sums.
- Independent $\Rightarrow \mathrm{Cov}(X, Y) = 0$ (because $E(XY) = E(X)E(Y)$). $\mathrm{Cov} = 0$ does **not** imply independence.

Uncorrelated but dependent (slides). $X = -1, 0, 1$ each with probability $1/3$ and $Y = X^2$. $E(X) = 0$, $E(Y) = 2/3$, $E(XY) = E(X^3) = 0$, so $\mathrm{Cov}(X, Y) = 0 - 0 \cdot (2/3) = 0$. Yet Y is a function of X. Covariance only detects linear association.

Three coins (slides). X = number of heads, Y = number of tails in 3 fair tosses. Since $Y = 3 - X$, $\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, 3 - X) = \mathrm{Cov}(X, 3) - \mathrm{Cov}(X, X) = -\mathrm{Var}(X) = -3(1/2)(1/2) = -3/4$.

Practice set, problem 16. $\mathrm{Cov}(4X - Y, X + Y) = 4\,\mathrm{Cov}(X, X) + 4\,\mathrm{Cov}(X, Y) - \mathrm{Cov}(Y, X) - \mathrm{Cov}(Y, Y) = 4\,\mathrm{Var}(X) + 3\,\mathrm{Cov}(X, Y) - \mathrm{Var}(Y)$: option (C).

Practice set, problem 24: socks. 8 left socks are matched at random with the 8 right socks (a random permutation). $I_{red} = 1$ if the red pair is matched correctly, I_{blue} likewise. P(red correct) = 1/8; P(red and blue both correct) = (1/8)(1/7) = 1/56. $\mathrm{Cov}(I_{red}, I_{blue}) = E(I_{red}I_{blue}) - E(I_{red})E(I_{blue}) = 1/56 - 1/64 = 1/448 \approx 0.00223$, in (0.001, 0.0025]. Slightly positive: a correct red match removes one wrong option for blue. (For indicators, $E(I_A I_B) = P(A \cap B)$, so $\mathrm{Cov} = P(A \cap B) - P(A)P(B)$.)

HW9 problem 10 (stock prices), part 1. $E(X) = 1$, $E(Y) = 3$, $E(Z) = 3$; $E(X^2) = 2$, $E(Y^2) = 12$, $E(Z^2) = 12$; $E(XY) = 2$, $E(XZ) = 4$, $E(YZ) = 9$. Then $\mathrm{Var}(X) = 2 - 1 = 1$, $\mathrm{Var}(Y) = 12 - 9 = 3$, $\mathrm{Var}(Z) = 3$; $\mathrm{Cov}(X, Y) = 2 - 3 = -1$, $\mathrm{Cov}(X, Z) = 4 - 3 = 1$, $\mathrm{Cov}(Y, Z) = 9 - 9 = 0$.

Extra context beyond the slides: multinomial counts are negatively correlated, $\mathrm{Cov}(X_i, X_j) = -n\,p_i p_j$. In HW9 problem 4, $\mathrm{Cov}(X_A, X_V) = -4(0.35)(0.45) = -0.63$.

12.5 Variance of a sum

CONSEQUENCE OF BILINEARITY (WARNING-SIGN SLIDE).

$$\mathrm{Var}(\Sigma_{i=1}^{n} X_i) = \Sigma_{i=1}^{n} \mathrm{Var}(X_i) + 2\,\Sigma_{i<j} \mathrm{Cov}(X_i, X_j),$$

because $\mathrm{Var}(\Sigma X_i) = \mathrm{Cov}(\Sigma X_i, \Sigma X_j) = \Sigma_i \Sigma_j \mathrm{Cov}(X_i, X_j)$, and the diagonal terms are the variances. Two variables: $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y)$ and $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$. If the X_i are independent (or merely uncorrelated), all cross terms vanish and $\mathrm{Var}(\Sigma X_i) = \Sigma\,\mathrm{Var}(X_i)$. Means always add; variances add only without covariance.

VAR(2X) IS NOT VAR($X_1 + X_2$). From the board: $\mathrm{Var}(2X) = 4\,\mathrm{Var}(X)$, but for independent copies $\mathrm{Var}(X_1 + X_2) = 2\,\mathrm{Var}(X)$. Doubling one variable doubles its spread; adding two independent copies only adds their variances.

The Binomial variance $np(1 - p)$ is n copies of the Bernoulli variance $p(1 - p)$, and the average of n independent copies has $\mathrm{Var}(\bar{X}) = \sigma^2/n$ (used in section 14).

HW10 problem 7. $X_0, \ldots, X_3$ independent with variance σ^2; $Y_1 = X_1 + 3X_0$, $Y_2 = X_2 + 2X_1$, $Y_3 = X_3 - X_2$. First simplify the sum: $Y_1 + Y_2 + Y_3 = 3X_0 + 3X_1 + X_3$ (the X_2 terms cancel). Then Var $= 9\sigma^2 + 9\sigma^2 + \sigma^2 = 19\sigma^2$: C = 19, in [15, 20), option (D). Adding $\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + \mathrm{Var}(Y_3) = 10 + 5 + 2 = 17$ would be wrong, because the Y's share X's and are correlated.

HW9 problem 9: symmetrization. X' and X'' are independent copies of X and $E[(X' - X'')^2] = 20$. Since $E(X' - X'') = 0$, $E[(X' - X'')^2] = \mathrm{Var}(X' - X'') = \mathrm{Var}(X) + \mathrm{Var}(X) = 2\mathrm{Var}(X)$. So $\mathrm{Var}(X) = 10$.

HW9 problem 10, part 3: the portfolio. $R = X + 2Y + Z$. $\mathrm{Var}(R) = \mathrm{Var}(X) + 4\mathrm{Var}(Y) + \mathrm{Var}(Z) + 2[2\mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z) + 2\mathrm{Cov}(Y, Z)] = 1 + 12 + 3 + 2[-2 + 1 + 0] = 14$, so $\sigma_R = \sqrt{14} \approx 3.7417$. The negative covariance between X and Y reduces the risk; that is diversification.

12.6 Correlation

CORRELATION COEFFICIENT (SLIDES). Assume $\mathrm{Var}(X) > 0$ and $\mathrm{Var}(Y) > 0$.

$$\rho(X, Y) = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{(\mathrm{Var}(X)\ \mathrm{Var}(Y))}}.$$

Fact 1: $-1 \le \rho \le 1$. Fact 2: if $\rho = -1$ then $Y = -aX + b$ with $a > 0$ (and $\rho = 1$ means $Y = aX + b$, $a > 0$): a perfect linear relationship. Fact 3: $\rho(aX + b, Y) = \rho(X, Y)$ for $a > 0$ (the sign flips if $a < 0$). X and Y are called uncorrelated when $\rho = 0$; that is weaker than independent. ρ measures the strength and direction of a *linear* relationship.

Why $|\rho| \le 1$ (in class). Standardize: $U = (X - EX)/\sqrt{\mathrm{Var}(X)}$ and $V = (Y - EY)/\sqrt{\mathrm{Var}(Y)}$, so $E(U) = E(V) = 0$, $E(U^2) = E(V^2) = 1$ and $\rho = E(UV)$. For any number t, $0 \le E[(U + tV)^2] = 1 + 2t\rho + t^2$. Taking $t = -\rho$ gives $0 \le 1 - \rho^2$, so $-1 \le \rho \le 1$. If $\rho = -1$ the choice $t = 1$ gives $E[(U + V)^2] = 0$, so $U = -V$ always, which unwinds to $Y = -aX + b$ with $a = \sqrt{(\mathrm{Var}\ Y/\mathrm{Var}\ X)} > 0$: Fact 2. Fact 3 follows from $\mathrm{Cov}(aX + b, Y) = a\,\mathrm{Cov}(X, Y)$ and $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

HW9 problem 10, parts 1–2. $\rho(X, Y) = -1/(\sqrt{1}\ \sqrt{3}) = -1/\sqrt{3} \approx -0.5774$. $\rho(Y, Z) = 0/(\sqrt{3}\ \sqrt{3}) = 0$: Y and Z are uncorrelated (from the given moments you cannot tell whether they are independent).

12.7 Check yourself

1. Two balls without replacement from 5 white and 8 red; X_1, X_2 indicate "red" on draws 1 and 2. Give $P(X_1 = 1, X_2 = 1)$ and $P(X_2 = 1)$, and decide whether X_1, X_2 are independent.

$(8/13)(7/12) = 14/39$; $8/13$; not independent since $(8/13)^2 = 64/169 \neq 14/39$.

2. $p(x, y) = (x^2 + y^2)/140$ for $x = 1, 2, 3$, $y = -x, \ldots, x$. Independent? [PRACTICE 13]

No: the support is not a rectangle and the formula is not a product.

3. Dogs arrive as Poisson(10) per day; each is "crazy" with probability 0.3 independently. Distribution of the number of crazy dogs? P(no crazy dogs today)?

Poisson(3), independent of the Poisson(7) normal dogs; $e^{-3} \approx 0.0498$.

4. X ~ Binomial(5, 2/3) and Y ~ Poisson(2) independent; Z = XY. Find E(Z) and Var(Z). [HW9 P8]

$E(Z) = 20/3 \approx 6.67$; $E(Z^2) = (110/9)(6) = 220/3$; $\mathrm{Var}(Z) = 260/9 \approx 28.89$.

5. Roll a fair four-sided die 20 times; Y = maximum. Find P(Y = 3). [HW9 P6]

$(3/4)^{20} - (1/2)^{20} \approx 0.00317$.

6. Toss a fair coin 3 times; X = heads, Y = tails. Cov(X, Y)? ρ(X, Y)?

$-\text{Var(X)} = -3/4$; $\rho = (-3/4)/(\sqrt{(3/4)}\sqrt{(3/4)}) = -1$ (Y is exactly 3 − X).

7. Express Cov(4X − Y, X + Y) in terms of Var(X), Var(Y), Cov(X, Y). Then Cov(2X + 3, 5Y − 1)? [Practice 16]

4Var(X) + 3Cov(X, Y) − Var(Y). 10 Cov(X, Y) (constants drop out).

8. 8 pairs of socks randomly re-paired left with right. Cov(I_{red}, I_{blue})? [Practice 24]

$1/56 - 1/64 = 1/448 \approx 0.00223$.

9. X_0, ..., X_3 independent with variance σ^2; $Y_1 = X_1 + 3X_0$, $Y_2 = X_2 + 2X_1$, $Y_3 = X_3 - X_2$. Var($Y_1 + Y_2 + Y_3$)? [HW10 P7]

$\text{Var}(3X_0 + 3X_1 + X_3) = 19\sigma^2$.

10. X′, X″ independent copies of X with $E[(X' - X'')^2] = 20$. Var(X)? [HW9 P9]

2Var(X) = 20, so Var(X) = 10.

11. With E(X) = 1, E(Y) = 3, E(Z) = 3, $E(X^2) = 2$, $E(Y^2) = 12$, $E(Z^2) = 12$, E(XY) = 2, E(XZ) = 4, E(YZ) = 9: find ρ(X, Y), ρ(Y, Z) and SD(X + 2Y + Z). [HW9 P10]

$-1/\sqrt{3} \approx -0.5774$; 0; $\sqrt{14} \approx 3.7417$.

12. X and Y independent with Var(X) = 2, Var(Y) = 3. Var(2X − 3Y)? Var(X + Y)? Var(X − Y)? And if X_1, X_2 are independent copies of X, Var($2X_1$) versus Var($X_1 + X_2$)?

8 + 27 = 35; 5; 5. $\text{Var}(2X_1) = 8$ but $\text{Var}(X_1 + X_2) = 4$.

13. X uniform on (0, 1), $Y = X^2$. Cov(X, Y)? Are they independent?

$E(XY) = E(X^3) = 1/4$; $E(X)E(Y) = (1/2)(1/3) = 1/6$; Cov = 1/12. Not independent (Y is a function of X); $\rho \approx 0.968$.

14. Two machines fail after independent Exponential times with means 10 and 5 hours. Distribution of the time until the first failure? P(the first failure is after 2 hours)?

Rates 0.1 and 0.2, so the minimum is Exponential(0.3) with mean 3.33 hours; $P(\min > 2) = e^{-0.6} \approx 0.5488$.

Deck 8 13. The Normal distribution

The idea. The Normal (Gaussian, "bell curve") is the distribution of standardized test scores, blood pressure and beta-carotene levels, the driving component of the Black–Scholes model in finance, and, by the central limit theorem (section 14), the approximate distribution of any sum or average of many independent terms. Every computation reduces to the standard Normal Z through standardization, and probabilities for Z come from its cdf Φ, which has no formula but is tabulated.

13.1 Definition, standardization and the table

Normal(μ, σ^2). $X \sim N(\mu, \sigma^2)$ if its density is

$$f(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}}\, e^{-(x-\mu)^2/(2\sigma^2)},\ -\infty < x < \infty,\ \text{with } E(X) = \mu,\ \mathrm{Var}(X) = \sigma^2,\ M_X(t) = e^{\mu t + \sigma^2 t^2/2}.$$

Note 3 (slides): if $a \neq 0$ then $aX + b \sim N(a\mu + b, a^2\sigma^2)$. N(0, 1) is the **standard Normal**; its cdf is $\Phi(x) = P(Z \le x) = \int_{-\infty}^{x} e^{-y^2/2}/\sqrt{(2\pi)}\, dy$, which cannot be written with elementary functions.

Standardize. If $X \sim N(\mu, \sigma^2)$ then $Z = (X - \mu)/\sigma \sim N(0, 1)$, so every $N(\mu, \sigma^2)$ computation reduces to N(0, 1):

$$P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right),\ P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),\ \Phi(-x) = 1 - \Phi(x).$$

For $x > 0$, $\Phi(x)$ is given in Table 5.1 of the textbook (the slide reads $P(Z \le 2.23) = 0.9871$ off the table); for $x < 0$ use the symmetry. The 68–95–99.7 rule: $P(|X - \mu| < \sigma) \approx 0.68$, $P(|X - \mu| < 2\sigma) \approx 0.95$, $P(|X - \mu| < 3\sigma) \approx 0.997$.

Note 2: why the constant is $1/\sqrt{(2\pi\sigma^2)}$ (in class). Let $I = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$. Then $I^2 = \iint e^{-(x^2+y^2)/2}\, dx\, dy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}\, r\, dr\, d\theta = \int_0^{2\pi} 1\, d\theta = 2\pi$ in polar coordinates, so $I = \sqrt{(2\pi)}$, and the substitution $x = \mu + \sigma z$ brings in the σ. Why Z has mean 0: $z\varphi(z)$ is odd. Why variance 1: integrate by parts, or use the MGF below.

z	0	0.25	0.5	0.75	1	1.25	1.5	1.645	1.75	1.96	2	2.25	2.5	3
$\Phi(z)$	0.5000	0.5987	0.6915	0.7734	0.8413	0.8944	0.9332	0.9500	0.9599	0.9750	0.9772	0.9878	0.9938	0.9987

Example. Heights are N(70, 9) inches ($\sigma = 3$). $P(X > 76) = 1 - \Phi(2) = 0.0228$. $P(67 < X < 73) = \Phi(1) - \Phi(-1) = 0.8413 - 0.1587 = 0.6826$. The 90th percentile is $\mu + 1.2816\sigma \approx 73.8$ inches, since $\Phi(1.2816) = 0.90$.

Variance or standard deviation?. $N(\mu, \sigma^2)$ is written with the variance. N(100, 225) has $\sigma = 15$. If a problem says "mean 2 and variance 4", divide by 2, not 4, when standardizing.

13.2 The MGF, the moments, and sums of Normals

Note 4 (in class). $M_X(t) = \frac{1}{\sigma\sqrt{(2\pi)}} \int e^{tx}\, e^{-(x-\mu)^2/(2\sigma^2)}\, dx$. Substitute $x = \mu + \sigma z$, $dx = \sigma\, dz$: $M_X(t) = \frac{1}{\sqrt{(2\pi)}} \int \exp(\mu t + \sigma t z - z^2/2)\, dz$. Complete the square, $\sigma t z - z^2/2 = -(z - \sigma t)^2/2 + \sigma^2 t^2/2$:

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2} \cdot \frac{1}{\sqrt{(2\pi)}} \int e^{-(z-\sigma t)^2/2}\, dz = e^{\mu t + \sigma^2 t^2/2},$$

because the remaining integrand is the density of a Normal shifted by σt and integrates to 1. For Z itself, $M_Z(t) = e^{t^2/2}$.

Note 5 (in class). $M'_X(t) = (\mu + \sigma^2 t)\, M_X(t)$, so $E(X) = M'(0) = \mu$. $M''_X(t) = \sigma^2 M_X(t) + (\mu + \sigma^2 t)^2 M_X(t)$, so $E(X^2) = \sigma^2 + \mu^2$ and $\mathrm{Var}(X) = \sigma^2$. (The slides suggest also computing $\int x\, f(x)\, dx$ and $\int x^2\, f(x)\, dx$ directly at home.) From M_Z, $E(Z^4) = 3$, so $\mathrm{Var}(Z^2) = 2$.

Note 3 via the MGF. $M_{aX+b}(t) = e^{bt} M_X(at) = e^{bt + a\mu t + a^2\sigma^2 t^2/2} = e^{(a\mu+b)t + (a^2\sigma^2)t^2/2}$, exactly the MGF of $N(a\mu + b,\, a^2\sigma^2)$. Standardizing is the case $a = 1/\sigma$, $b = -\mu/\sigma$.

> SUMS OF INDEPENDENT NORMALS ARE NORMAL (IN CLASS). If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent, $M_{X+Y}(t) = M_X(t) M_Y(t) = e^{(\mu_X+\mu_Y)t + (\sigma_X^2+\sigma_Y^2)t^2/2}$, so $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. More generally $\Sigma\, a_i X_i \sim N(\Sigma\, a_i \mu_i, \Sigma\, a_i^2 \sigma_i^2)$, and the average of n independent $N(\mu, \sigma^2)$ is $N(\mu, \sigma^2/n)$.

Example. $X \sim N(2, 4)$ and $Y \sim N(1, 5)$ independent. $X - Y \sim N(1, 9)$, so $P(X > Y) = P(X - Y > 0) = 1 - \Phi(-1/3) = \Phi(1/3) \approx 0.6306$.

13.3 E|Z| (HW10 problem 11)

$$E|Z| = \int_{-\infty}^{\infty} |z|\, \varphi(z)\, dz = 2 \int_0^{\infty} z\, \frac{e^{-z^2/2}}{\sqrt{(2\pi)}}\, dz = \frac{2}{\sqrt{(2\pi)}} \left[-e^{-z^2/2}\right]_0^{\infty} = \frac{2}{\sqrt{(2\pi)}} = \sqrt{(2/\pi)} \approx 0.7979.$$

(The substitution $u = z^2/2$ does the integral.) So the average distance of a standard Normal from 0 is about 0.8, less than its standard deviation 1, as it should be: $E|Z| \le \sqrt{E(Z^2)} = 1$. HW10 problem 11 then gives $E[|X|(Y+1)^2] = \sqrt{(2/\pi)} \cdot 5 \approx 3.989$ (section 12.1).

13.4 Check yourself

1. For Z standard Normal: $P(Z \le 2.23)$, $P(Z > 1.5)$, $P(-1 < Z < 2)$, $P(|Z| > 2)$.

0.9871 (table); $1 - 0.9332 = 0.0668$; $0.9772 - 0.1587 = 0.8185$; $2(0.0228) = 0.0456$.

2. IQ scores are N(100, 225). P(score > 130)? P(85 < score < 115)?

$\sigma = 15$: $1 - \Phi(2) = 0.0228$; $\Phi(1) - \Phi(-1) = 0.6826$.

3. $X \sim N(0, 1)$, $Y \sim$ Exponential(1), independent. $E[|X|(Y+1)^2]$? [HW10 P11]

$\sqrt{(2/\pi)} \times 5 \approx 3.989$.

4. $X \sim N(2, 4)$, $Y \sim N(1, 5)$, independent. $P(X > Y)$?

$X - Y \sim N(1, 9)$: $\Phi(1/3) \approx 0.6306$.

5. A random variable has MGF $e^{3t + 8t^2}$. Name its distribution, mean and variance. Then $P(X > 7)$?

Normal with $\mu = 3$ and $\sigma^2/2 = 8$, so $\sigma^2 = 16$, $\sigma = 4$. $P(X > 7) = 1 - \Phi(1) = 0.1587$.

6. If $X \sim N(\mu, \sigma^2)$, what is the distribution of $3X - 2$? Of $(X - \mu)/\sigma$?

$N(3\mu - 2, 9\sigma^2)$; $N(0, 1)$.

7. $E(Z^2)$, $E(Z^4)$ and $\mathrm{Var}(Z^2)$ for Z standard Normal?

1; 3; 2.

8. The average of 16 independent $N(50, 64)$ measurements: distribution, and P(the average exceeds 53)?

$N(50, 64/16) = N(50, 4)$, $\sigma = 2$: $P = 1 - \Phi(1.5) = 0.0668$.

Deck 8 14. Limit theorems: the CLT, the law of large numbers, and the inequalities

Where this sits. These are the last slides of deck 8, and sections 6–7 of the instructor's Lectures 5–8 study guide mark them as warning-sign material. (The syllabus had listed them under deck 9, but deck 9 itself is jointly continuous random variables, which is not on this midterm.)

The idea. Why does a sample of about 1100 people predict an election? How long must a simulation run? The answers come from limit theory. The law of large numbers says the average of many independent observations settles at the mean. The central limit theorem says the fluctuations around the mean are Normal, with size $\sigma/\sqrt{n}$, whatever the original distribution; it is why the Normal is everywhere and how we approximate Binomial probabilities. Markov's and Chebyshev's inequalities give guaranteed bounds from a mean alone or a mean and a variance, and Chebyshev is what proves the law of large numbers.

14.1 The central limit theorem

CLT (warning-sign slide). If $X_1, X_2, \ldots$ are independent, identically distributed with mean μ and variance $0 < \sigma^2 < \infty$, then the distribution of

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard Normal as $n \to \infty$: $P(Z_n \le a) \to \Phi(a)$ for every a. Punchline: for large n,

$$S_n = X_1 + \cdots + X_n \approx N(n\mu, n\sigma^2) \text{ and } \bar{S}_n = S_n/n \approx N(\mu, \sigma^2/n).$$

The numerator subtracts the mean of the sum; the denominator is the standard deviation of the sum. (Annotation on the slide: "mean of each X_i already subtracted off, divided by σ".)

Idea of the proof (in class). Assume the MGF of X_i exists. Let $U_i = (X_i - \mu)/\sigma$, so $E(U_i) = 0$, $E(U_i^2) = 1$ and $Z_n = (U_1 + \cdots + U_n)/\sqrt{n}$. A second-order Taylor expansion at $s = 0$ gives $M_U(s) = 1 + E(U)s + E(U^2)s^2/2 + o(s^2) = 1 + s^2/2 + o(s^2)$. By independence, $M_{Z_n}(t) = [M_U(t/\sqrt{n})]^n = [1 + t^2/(2n) + o(1/n)]^n \to e^{t^2/2}$, the MGF of N(0, 1), using $(1 + x/n)^n \to e^x$. Fact: if the MGFs converge, the cdfs converge.

Example 1 (slides): a random walk. $X_i = \pm 1$ with probability 1/2 each (a stock price moving up or down by 1 each day). $P(S_{50} > 10)$?

1. One step: $E(X_i) = 0$, $E(X_i^2) = 1$, so $Var(X_i) = 1$.
2. By independence, $E(S_{50}) = 0$ and $Var(S_{50}) = 50$; by the CLT, $S_{50} \approx N(0, 50)$.
3. $P(S_{50} > 10) \approx P(Z > 10/\sqrt{50}) = P(Z > 1.41) = 1 - \Phi(1.41) \approx 1 - 0.9207 = 0.079$.

Extra context beyond the slides: S_{50} only takes even values, so "$S_{50} > 10$" means $S_{50} \ge 12$ and a continuity correction uses $P(W > 11) = 1 - \Phi(11/\sqrt{50}) = 1 - \Phi(1.56) \approx 0.060$, very close to the exact 0.0595. The uncorrected 0.079 is what the lecture computed; either is acceptable as a CLT approximation unless the problem asks for the correction.

Example 3 (slides): the astronomer. Measurements are iid with mean d (the true distance) and variance 4. How many are needed to be "reasonably sure" (say 95%) that the average is within ± 0.5 light-year of d? The average $\bar{X}_n \approx N(d, 4/n)$, so $P(|\bar{X}_n - d| \le 0.5) = P(|Z| \le 0.5/(2/\sqrt{n})) = P(|Z| \le 0.25\sqrt{n})$. For this to be 0.95 we need $0.25\sqrt{n} \ge 1.96$, i.e. $n \ge (7.84)^2 \approx 61.5$: about 62 measurements. (Chebyshev, in 14.4, gives the distribution-free but much more conservative $n \ge 320$.)

HW10 problem 1. $X_i = 1$ or 2 with probabilities 1/3, 2/3. $\mu = 1/3 + 4/3 = 5/3$; $E(X_i^2) = 1/3 + 8/3 = 3$, so $\sigma^2 = 3 - 25/9 = 2/9$ and $\sigma = \sqrt{2}/3$. The CLT says the distribution of

$$\frac{X_1 + \cdots + X_n - 5n/3}{\sqrt{(2n/9)}} = \frac{X_1 + \cdots + X_n - 5n/3}{(\sqrt{2}/3)\sqrt{n}}$$

tends to the standard Normal. The blanks are $n\mu$ and $\sigma\sqrt{n}$.

14.2 Normal approximation to the Binomial and the continuity correction

BINOMIAL ≈ NORMAL. A Binomial(n, p) is a sum of n Bernoullis, so by the CLT $X \approx W \sim N(np, np(1-p))$ when n is large. Subtle point (slides): the Binomial is discrete and the Normal is continuous, so use the **continuity correction**:

$$P(X = k) \approx P(k - 0.5 < W < k + 0.5),\ P(X \le k) \approx P(W < k + 0.5),\ P(X < k) \approx P(W < k - 0.5),$$

and likewise $P(X \ge k) \approx P(W > k - 0.5)$, $P(X > k) \approx P(W > k + 0.5)$. Move each integer cutoff by half a unit in the direction that captures the bars of the discrete distribution.

The beer example (slides). Eva's PR firm tests 100 affirmed Bond-beer drinkers blind; X = number who prefer VP beer. In class the "worst case" $p = 1/2$ was used (most people cannot tell non-specialty beers apart), so $X \sim \text{Bin}(100, 1/2)$: $\mu = np = 50$, $\sigma^2 = np(1-p) = 25$, $\sigma = 5$, and $W \sim N(50, 25)$.

1. $P(X > 40) \approx P(W > 40.5) = P(Z > (40.5 - 50)/5) = P(Z > -1.9) = \Phi(1.9) \approx 0.9713$.
2. $P(40 \le X \le 60) \approx P(39.5 < W < 60.5) = P(-2.1 < Z < 2.1) = \Phi(2.1) - \Phi(-2.1) = 2\Phi(2.1) - 1 \approx 0.9643$.

(Exact Binomial values: 0.9716 and 0.9648. Without the correction the answers would be 0.9772 and 0.9545, noticeably worse.) Rule of thumb: the approximation is good when $np(1-p)$ is at least about 10.

Another example. 100 fair coin tosses; P(at least 60 heads) $\approx P(W > 59.5) = 1 - \Phi(9.5/5) = 1 - \Phi(1.9) \approx 0.0287$ (exact 0.0284).

14.3 The weak law of large numbers

WEAK LLN (WARNING-SIGN SLIDE). If $X_1, X_2, \ldots$ are independent, identically distributed with mean μ, then for every $\varepsilon > 0$,

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \ge \varepsilon\right) \to 0 \text{ as } n \to \infty.$$

Punchline: the sample mean gets closer and closer to the population mean as the sample size grows.

Exercise on the slide. With $X_i = I_{A_i}$ the indicator that event A happens on trial i (so $\mu = P(A)$), the law says the observed proportion of trials on which A occurs converges to P(A). That is the frequency interpretation of probability, now a theorem, and the reason the London-bomb and Gretzky tables match the Poisson so well.

14.4 Markov's and Chebyshev's inequalities

Markov's inequality (warning-sign slide). If X is a non-negative random variable, then for any $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof (in class): since $X \geq 0$ and $a > 0$, $X \geq a \cdot 1_{\{X \geq a\}}$ always (if $X \geq a$ the right side is $a \leq X$; otherwise it is $0 \leq X$). Take expectations: $E(X) \geq a\, P(X \geq a)$; divide by a.

Chebyshev's inequality (warning-sign slide). If X has mean μ and variance σ^2, then for any $a > 0$,

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}, \text{ equivalently } P(|X - \mu| < a) \geq 1 - \frac{\sigma^2}{a^2}.$$

Proof (in class): $Y = (X - \mu)^2$ is a non-negative random variable, so Markov gives $P(Y \geq a^2) \leq E(Y)/a^2 = \mathrm{Var}(X)/a^2$; and $\{Y \geq a^2\}$ is the same event as $\{|X - \mu| \geq a\}$. With $a = k\sigma$: at most $1/k^2$ of any distribution lies k or more standard deviations from its mean (compare the Normal's 0.0456 at $k = 2$).

Example 2a (slides): the factory. Weekly production X has mean 50.

1. $P(X > 75) \leq P(X \geq 75) \leq 50/75 = 2/3$ by Markov (production is non-negative).
2. If also $\mathrm{Var}(X) = 25$: $P(40 \leq X \leq 60) = P(|X - 50| \leq 10) = 1 - P(|X - 50| > 10) \geq 1 - P(|X - 50| \geq 10) \geq 1 - 25/100 = 3/4$ by Chebyshev.

Nothing was assumed about the shape of the distribution; that is the point, and also why the bounds are loose.

HW10 problem 2. $X \sim \mathrm{Poisson}(100)$, so $E(X) = \mathrm{Var}(X) = 100$. (a) Markov: $P(X > 120) \leq P(X \geq 120) \leq 100/120 \approx 0.8333$ (or, using the integer values, $\leq 100/121 \approx 0.8264$; both are valid bounds). (b) Chebyshev: $\{X > 120\}$ is contained in $\{|X - 100| \geq 20\}$, so $P(X > 120) \leq 100/400 = 0.25$. The true probability is about 0.023.

The astronomer by Chebyshev. $P(|X_n - d| \geq 0.5) \leq \mathrm{Var}(X_n)/0.25 = (4/n)/0.25 = 16/n$. For this to be at most 0.05 we need $n \geq 320$, versus about 62 from the CLT: Chebyshev holds for every distribution but pays for it in sharpness.

One-sided tails. Chebyshev is two-sided. To bound $P(X > \mu + a)$, note that the one-sided event is a subset of $\{|X - \mu| \geq a\}$ and use the same σ^2/a^2. To bound "between": complement, as in the factory example.

14.5 Proof of the weak law (how Chebyshev does it)

Proof (slides, assuming finite variance σ^2). Let $X_n = (X_1 + \cdots + X_n)/n$. By linearity, $E(X_n) = (1/n)\,\Sigma\, E(X_i) = \mu$. By independence, $\mathrm{Var}(X_n) = (1/n^2)\,\Sigma\, \mathrm{Var}(X_i) = \sigma^2/n$. For any fixed $\varepsilon > 0$, Chebyshev applied to X_n gives

$$P(|X_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(X_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

Practice set, problem 18. The bound used in the proof is $\sigma^2/(\varepsilon^2 n)$: option (C). Options with n in the numerator get worse with more data, which cannot be right.

KEEP THE THREE LIMIT FACTS APART. Chebyshev is a *bound* (never an approximation). The weak law says $\bar{X} \to \mu$. The CLT says the shape of $\bar{X}$'s fluctuations is Normal with standard deviation $\sigma/\sqrt{n}$, and that $S_n \approx N(n\mu, n\sigma^2)$. "Fill in the blanks so that $(S_n - ____)/____$ tends to the standard Normal" wants $n\mu$ and $\sigma\sqrt{n}$.

14.6 Check yourself

1. $X_i = 1$ or 2 with probabilities 1/3, 2/3, independent. Fill in: $(X_1 + \cdots + X_n - ____)/____$ tends to $N(0, 1)$. [HW10 P1]

$5n/3$ and $\sqrt{(2n/9)} = (\sqrt{2}/3)\sqrt{n}$.

2. $X \sim$ Poisson(100). Bound $P(X > 120)$ using Markov, then using Chebyshev. [HW10 P2]

Markov: $\leq 100/120 \approx 0.833$. Chebyshev: $\leq 100/20^2 = 0.25$.

3. In the proof of the weak law, Chebyshev bounds $P(|\bar{X}_n - \mu| \geq \varepsilon)$ by what? [PRACTICE 18]

$\sigma^2/(n\varepsilon^2)$.

4. Weekly production has mean 50 and variance 25. Bound $P(X > 75)$, and give a lower bound for $P(40 \leq X \leq 60)$.

$\leq 2/3$ (Markov); $\geq 3/4$ (Chebyshev).

5. A random walk takes 100 steps of ± 1 (probability 1/2 each). Approximate $P(S_{100} > 20)$ with the CLT.

$S_{100} \approx N(0, 100)$: $1 - \Phi(20/10) = 1 - \Phi(2) \approx 0.0228$ (with the continuity correction, $1 - \Phi(2.1) \approx 0.018$; exact 0.0176).

6. 100 blind tasters, each preferring VP beer with probability 1/2. Approximate $P(X > 40)$ and $P(40 \leq X \leq 60)$ with the continuity correction.

$\Phi(1.9) \approx 0.9713$; $2\Phi(2.1) - 1 \approx 0.9643$.

7. Measurements are iid with variance 4. How many are needed so that the sample mean is within 0.5 of the true value with probability 0.95, by the CLT? By Chebyshev?

$0.25\sqrt{n} \geq 1.96$ gives $n \approx 62$; $16/n \leq 0.05$ gives $n \geq 320$.

8. 100 fair coin tosses. Approximate P(at least 60 heads) with the continuity correction.

$1 - \Phi(9.5/5) = 1 - \Phi(1.9) \approx 0.0287$.

9. The sample mean of 36 independent observations with $\sigma = 12$. $\mathrm{Var}(\bar{X})$? Approximate $P(|\bar{X} - \mu| < 4)$?

$144/36 = 4$, SD 2; $P(|Z| < 2) \approx 0.9545$.

10. A non-negative X has mean 10. Upper bound on P(X ≥ 50)? If also Var(X) = 16, lower bound on P(2 < X < 18)?

Markov: 0.2. Chebyshev with a = 8: $\geq 1 - 16/64 = 0.75$.

15. Exam strategy

15.1 Recognize the distribution from the story

Most questions are "identify the model, write its parameters, use one formula". Decide the model before touching the calculator.

The story says	Model	Parameters	First formula to write
n independent yes/no trials, same p, count the successes	Binomial(n, p)	n, p (p may itself need a total-probability step)	$\binom{n}{k}p^k(1-p)^{n-k}$; mean np, variance $np(1-p)$
Many trials, tiny p, "use the Poisson approximation"	Poisson(λ)	$\lambda = np$	$e^{-\lambda}\lambda^k/k!$
Events at a constant rate over a time window	Poisson process	λ = rate × window length (same units!)	$N(t) \sim \text{Poisson}(\lambda t)$; wait for the first event ~ Exponential(rate)
Trials until the first success	Geometric(p)	p	$(1-p)^{k-1}p$; mean $1/p$; $P(X > k) = (1-p)^k$
A random point in an interval; "equally likely anywhere"	Uniform(a, b)	a, b	length/(b − a); mean (a + b)/2; variance $(b-a)^2/12$
Waiting time, lifetime, "memoryless", "mean lifetime"	Exponential(λ)	$\lambda = 1/\text{mean}$	$P(X > x) = e^{-\lambda x}$
Several categories, fixed n, "exactly n_1 of type 1, n_2 of type 2 …"	Multinomial	n; p_1, …, p_r	$n!/(n_1! \cdot \cdots \cdot n_r!)\ \Pi\ p_i^{n_i}$
Draws without replacement, counts of each type	Counting (hypergeometric)	population sizes, draws	$\Pi\binom{N_i}{k_i} / \binom{N}{n}$
"Number of … that …" with no clean distribution	Indicators	one event per item	$E(X) = \Sigma\ P(A_i)$
Bell-shaped measurement; sums or averages of many terms	Normal / CLT	μ, σ^2 (of the sum: nμ, $n\sigma^2$)	Standardize, read Φ; ±0.5 continuity correction for counts
Only a mean (and maybe a variance) is known; "what can be said about …"	Markov / Chebyshev	E(X); σ^2	$E(X)/a$; σ^2/a^2

15.2 Four questions before you compute

1. **What is the random variable, and what values can it take?** Write "X = number of …" or "X = time until …" and its range. Half of the practice set is answered correctly by this alone.
2. **Discrete or continuous?** Discrete: sums, endpoints matter, P(X = a) can be positive. Continuous: integrals, endpoints never matter, P(X = a) = 0.
3. **Independent or not?** Independence lets you multiply probabilities and MGFs and add variances. Means add regardless. If the problem says "given", "conditional on", or describes one variable in terms of another, there is dependence and you need total probability or a joint pmf.
4. **Is the complement easier?** "At least one", "more than two", "at most n − 1": compute the small side and subtract from 1.

15.3 Multiple-choice tactics

- **Interval answers:** compute to four decimals, then check whether the endpoints are included: (0.25, 0.40] includes 0.40, [0.5, 0.65) excludes 0.65. When your number lands within 0.005 of an endpoint, recompute with more precision before choosing.
- **Formula answers:** test the options at an easy value ($p = 0$, $n = 1$, $t = 0$). An MGF option must equal 1 at $t = 0$. A pmf option must give probabilities between 0 and 1 at the smallest and largest values. Options that *add* MGFs or multiply variances by a instead of a^2 are wrong on sight.
- **Sanity limits:** variances are never negative; a probability from a Poisson or Binomial with mean 2 should not be 0.9; E(number of successes) $\le n$; E(X) for an Exponential is $1/\lambda$, not λ; the median of an Exponential is below its mean.
- **Units:** rate per hour times minutes is the most common silent error. Convert first.
- **Time:** about 6 minutes per question for 12 questions in 75 minutes. Do the recognizable ones first; MGF and covariance algebra questions are usually fastest.

15.4 The traps

Trap	Correct thinking
Using P(positive | sick) when the person tested negative	Every term in Bayes must condition on the result actually observed.
Calling disjoint events independent	Disjoint with positive probabilities $\Rightarrow$ dependent (one rules out the other).
$P(1 \le X \le 2) = F(2) - F(1)$ for a discrete X	$F(2) - F(\text{just below } 1) = F(2) - F(0)$ for integer values.
Treating f(a) as a probability	f(a) is a density; probabilities are areas. f can exceed 1.
$E(X^2) = (E\ X)^2$	$E(X^2) = \text{Var}(X) + (E\ X)^2$; the plug-in rule for E g(X).
$\text{Var}(aX + b) = a\text{Var}(X) + b$	$a^2\text{Var}(X)$; the b vanishes. SD scales by $\lvert a\rvert$.
$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ always	Only if uncorrelated; in general add $2\text{Cov}(X, Y)$. Means add always.
$M_{3X-2Y}(t) = M_X(t)^3\, M_Y(t)^{-2}$	$M_X(3t)\, M_Y(-2t)$, and only for independent X, Y.
Poisson process over 2 hours at 4 per week with $\lambda = 4$	λt = rate $\times$ length in the rate's units.
Exponential "mean 2" used as $\lambda = 2$	$\lambda = 1/\text{mean} = 1/2$.
Forgetting the age in a memoryless question, or including it	$P(X > s + t \mid X > s) = e^{-\lambda t}$: the extra time only.
Dividing by 2! in a multinomial with two distinguishable groups of the same size	Categories are labeled; $n!/(n_1! \cdots n_r!)$ with no extra division.
Concluding independence from Cov = 0	Cov = 0 is weaker; check the joint pmf factors (and the support is a rectangle).
Adding $\text{Var}(Y_1) + \text{Var}(Y_2) + \text{Var}(Y_3)$ when the Y's share terms	Simplify the sum first, then apply the variance to independent pieces.
Using Chebyshev as an approximation	It is an upper bound; the true probability is usually far smaller.
CLT denominator σ or σn	$\sigma\sqrt{n}$ for the sum; $\sigma/\sqrt{n}$ for the average.
Normal approximation to a count without the ± 0.5	$P(X \le k) \approx P(W < k + 0.5)$; $P(X \ge k) \approx P(W > k - 0.5)$.

16. Practice exams

Part A is the practice set posted for Midterm 2, with worked answers. Part B is a new set in the same style. Do each part in 75 minutes with only a calculator, then read every solution, including the ones you got right. The real exam will not just be variants of these, so focus on *which concept* each question uses.

16.1 Part A: the posted practice problems

Every Part A answer below opens with **the one big concept** the question is testing, then works the arithmetic in full and flags the traps the distractors are built from. Read the concept line first — if you can name it before you compute, you can handle the exam's version of the same question.

A1. Two fair dice (red and blue). E = the sum is 7, F = the red die shows 4. E and F are: (A) disjoint, (B) independent, (C) neither. [§3]

Big concept: independence is a *numerical* condition, $P(E \cap F) = P(E)P(F)$ — not the same as disjoint. Work on the 36 equally likely (red, blue) pairs. E = "sum is 7" = {(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)}, so $P(E) = 6/36 = 1/6$. F = "red shows 4", so $P(F) = 6/36 = 1/6$. For $E \cap F$ the red die is pinned at 4, so the blue die must be 3: exactly one outcome, $P(E \cap F) = 1/36$. Compare: $P(E)P(F) = (1/6)(1/6) = 1/36$. They match, so E and F are independent. They are certainly *not* disjoint, since (4,3) is in both — disjoint would force $P(E \cap F) = 0$. **(B) independent.**
Why this works and when it breaks: a sum of 7 is the one total that is equally likely under every value of the red die (whatever red shows, exactly one blue value completes it), so learning red = 4 tells you nothing about the sum being 7. Change the target to "sum is 6" and that symmetry dies: $P(E) = 5/36$, $P(E \cap F) = P(\text{red 4, blue 2}) = 1/36$, but $P(E)P(F) = 5/216 \neq 1/36$ — neither disjoint nor independent, answer (C).

A2. John attends with probability 0.65; Joe attends with probability 0.8 if John does and 0.45 if John does not. P(exactly one attends) is in: (A) (0, 0.35], (B) (0.35, 0.45], (C) (0.45, 0.6], (D) (0.6, 0.8], (E) (0.8, 1]. [§2]

Big concept: law of total probability — split on the conditioning event, and read "exactly one" as a union of disjoint cases.
Joe's probability is *conditional* on John, so you cannot multiply marginals. Condition on John:

- John attends (0.65) and Joe does not ($1 - 0.8 = 0.2$): $0.65(0.2) = 0.13$.
- John does not attend (0.35) and Joe does (0.45): $0.35(0.45) = 0.1575$.

These two cases are disjoint and cover "exactly one", so add: $0.13 + 0.1575 = \mathbf{0.2875}$, which sits in (0, 0.35]. **(A).**
Trap: treating the two attendances as independent with P(Joe) = 0.8 gives $0.65(0.2) + 0.35(0.8) = 0.41$ — wrong, because 0.8 only applies in the branch where John attends. Always ask whether a given number is a marginal or a conditional.

A3. A rolls a fair 4-sided and a fair 6-sided die, aiming for sum 3; then B rolls them aiming for sum 4; they alternate until someone succeeds. P(A wins) is in: (A) [0, 0.25), (B) (0.25, 0.45], (C) (0.45, 0.52], (D) (0.52, 0.55], (E) none. [§3.5]

Big concept: a "first success wins" race — use the restart (renewal) argument instead of summing an infinite series by hand.

First get the per-turn success probabilities on a 4-sided × 6-sided roll (24 equally likely pairs). Sum 3: (1,2), (2,1) → $p_A = 2/24 = 1/12$. Sum 4: (1,3), (2,2), (3,1) → $p_B = 3/24 = 1/8$.

Now condition on the first full round. A wins immediately with probability p_A. If A fails and B fails — probability $(1 - p_A)(1 - p_B)$ — the situation is *identical* to the start, so A's win probability q satisfies

$$q = p_A + (1 - p_A)(1 - p_B)\, q \quad \Rightarrow \quad q = \frac{p_A}{1 - (1\text{-}p_A)(1\text{-}p_B)}.$$

Here $(11/12)(7/8) = 77/96$, so the denominator is $19/96$ and $q = (1/12)(96/19) = \mathbf{8/19 \approx 0.4211}$, in (0.25, 0.45]. **(B).**

Sanity check: A goes first (an advantage) but has the smaller per-turn probability ($1/12 < 1/8$), and the second effect wins, so the answer should land a bit below 1/2. The same formula is the geometric-series sum $\sum_{k\geq 0} [(1-p_A)(1-p_B)]^k\, p_A$.

A4. E, F, G, H, I independent with probabilities 0.2, 0.3, 0.1, 0.4, 0.5. P((E ∪ F^c ∪ G) ∩ H | I) is in: (A) (0, 0.1], (B) (0.1, 0.15], (C) (0.15, 0.25], (D) (0.25, 0.5], (E) (0.5, 1]. [§3.3]

Big concept: independence passes through complements, unions and intersections — so conditioning on an independent event changes nothing, and unions are handled by complementing.

Step 1 (drop the condition): I is independent of every event built from E, F, G, H, so $P((E \cup F^c \cup G) \cap H \mid I) = P((E \cup F^c \cup G) \cap H)$. The 0.5 is a decoy.

Step 2 (the union, via its complement): $P(E \cup F^c \cup G) = 1 - P(E^c \cap F \cap G^c) = 1 - (0.8)(0.3)(0.9) = 1 - 0.216 = 0.784$. Note the complement of F^c is F, with probability 0.3 — not 0.7.

Step 3 (the intersection): multiply by $P(H) = 0.4 \rightarrow 0.784(0.4) = \mathbf{0.3136}$, in (0.25, 0.5]. **(D).**

Why complement the union: inclusion–exclusion on three events needs seven terms; complementing turns it into one product. This "1 − (all fail)" move is the single most reused trick on independence problems.

A5. 40% of adults are addicted; the test is positive for 70% of addicted and 20% of non-addicted adults. A random adult tests negative. P(addicted) is in: (A) (0, 0.15], (B) (0.15, 0.3], (C) (0.3, 0.4], (D) (0.4, 0.45], (E) (0.45, 0.5]. [§2.2]

Big concept: Bayes' theorem, run on a *negative* result — so the relevant rates are the false-negative and true-negative rates.

Let A = addicted. Given: $P(A) = 0.4$ (the prior/base rate), $P(+ \mid A) = 0.7$ so $P(- \mid A) = 0.3$, and $P(+ \mid A^c) = 0.2$ so $P(- \mid A^c) = 0.8$. Flip the conditioning:

$$P(A \mid -) = \frac{P(A)P(- \mid A)}{P(A)P(- \mid A) + P(A^c)P(- \mid A^c)} = \frac{0.4(0.3)}{0.4(0.3) + 0.6(0.8)} = \frac{0.12}{0.12 + 0.48} = \frac{0.12}{0.60} = \mathbf{0.20},$$ in (0.15, 0.3]. **(B).**

Counting version (often faster): per 1000 adults, 400 addicted (120 test negative) and 600 not (480 test negative); 600 negatives total, 120 of them addicted → 120/600 = 0.2.

Sanity check: a negative result is evidence *against* addiction, so the answer must be below the 0.4 prior. The denominator is always P(evidence) — the total probability of the thing you actually observed.

A6. Each kidney matches with probability 0.1 independently; X = number tested until a match. E(X) is in: (A) (0, 1), (B) [1, 6), (C) [6, 12), (D) [12, 20), (E) [20, ∞). [§11.1]

Big concept: the Geometric distribution counts trials until the first success, and E(X) = 1/p.
Each kidney independently matches with p = 0.1, and you stop at the first match, so X ~ Geometric(0.1) in the "number of trials including the success" convention: $P(X = k) = (0.9)^{k-1}(0.1)$. Then E(X) = 1/p = **10**, in [6, 12). **(C).**
Also know: $Var(X) = (1 - p)/p^2 = 0.9/0.01 = 90$, SD ≈ 9.49 — a geometric is very spread out. And watch the convention: if the book counts *failures before* the first success, the mean is (1 − p)/p = 9, which would still land in (C) here but not in general.
Intuition: at a 1-in-10 success rate you expect to wait about 10 tries — the mean is the reciprocal of the rate. This is the discrete cousin of the Exponential's mean 1/λ.

A7. Policy holders: 50% standard (death probability 0.4), 40% preferred (0.2), 10% ultra-preferred (0.1). P(a random holder dies next year) is in: (A) (0, 0.1], (B) (0.1, 0.2], (C) (0.2, 0.25], (D) (0.25, 0.275], (E) (0.275, 1.0]. [§2.1]

Big concept: law of total probability as a weighted average over a partition of the population.
The three classes partition the policy holders, so weight each death probability by its class share:
P(dies) = 0.5(0.4) + 0.4(0.2) + 0.1(0.1) = 0.20 + 0.08 + 0.01 = **0.29**, which lies in (0.275, 1.0]. **(E).**
Checks that catch errors: the weights must sum to 1 (0.5 + 0.4 + 0.1 = 1 ✓), and the answer must land between the smallest and largest conditional probability (0.1 ≤ 0.29 ≤ 0.4 ✓). It sits near 0.4 because the highest-risk class carries half the weight.
Companion skill: the natural follow-up "given a holder died, P(standard)?" is the Bayes version — 0.20/0.29 ≈ 0.69 — reusing this same 0.29 as the denominator.

A8. The cdf of X is 0 below −2, 0.1 on [−2, −0.75), 0.3 on [−0.75, 0), 0.6 on [0, 1), 0.8 on [1, 2), 1 from 2 on. E(X) is in: (A) (−0.1, 0], (B) (0, 0.1], (C) (0.1, 0.2], (D) (0.2, 0.3], (E) (0.3, 0.4]. [§4.3, §6.1]

Big concept: for a discrete random variable the pmf is the set of *jump sizes* of the cdf; then E(X) = ∑ x p(x).
Read off the jumps of F at each point where it steps up:
at −2: 0.1 − 0 = 0.1 | at −0.75: 0.3 − 0.1 = 0.2 | at 0: 0.6 − 0.3 = 0.3 | at 1: 0.8 − 0.6 = 0.2 | at 2: 1 − 0.8 = 0.2.
They sum to 1 ✓. Now
E(X) = (−2)(0.1) + (−0.75)(0.2) + (0)(0.3) + (1)(0.2) + (2)(0.2) = −0.2 − 0.15 + 0 + 0.2 + 0.4 = **0.25**, in (0.2, 0.3]. **(D).**
Key facts about a cdf: it is right-continuous and the mass sits at the *left* end of each interval of constancy — the value 0.6 "on [0, 1)" means the jump happens exactly at 0, so P(X = 0) = 0.3 and P(X ≤ 0) = 0.6. If asked for Var, square the same x values: $E(X^2) = 0.4 + 0.1125 + 0 + 0.2 + 0.8 = 1.5125$, so Var = 1.5125 − 0.0625 = 1.45.

A9. 40 employees vaccinated with a 70% effective vaccine; X = number who get the flu. (A) E = 28, SD = 8.4; (B) E = 28, SD = 2.90; (C) E = 12, SD = 8.4; (D) E = 12, SD = 2.90; (E) E = 3500, SD = 8.4. [§8.3]

Big concept: Binomial mean and variance — E = np, Var = np(1 − p), SD = $\sqrt{}$(np(1−p)) — after defining "success" to match the question.
X counts employees who *get the flu*, which happens when the vaccine fails, so the success probability for X is p = 1 − 0.7 = 0.3, and X ~ Binomial(40, 0.3) (fixed n, two outcomes, independent, constant p).
E(X) = 40(0.3) = **12**; Var(X) = 40(0.3)(0.7) = **8.4**; SD = $\sqrt{8.4} \approx$ **2.90**. **(D).**
The two traps the distractors are built from: (i) using p = 0.7 gives E = 28 — that is the count of *protected* employees, answering the wrong question; (ii) reporting 8.4 as the SD instead of the variance. Note SD < variance here only because 8.4 > 1; always take the square root last.
Rule of thumb: Var = np(1−p) is largest at p = 0.5 and shrinks toward 0 as p approaches 0 or 1 — extreme probabilities make outcomes predictable.

A10. Same setting; vaccines cost $70 each and each sick employee costs $1000. Var(C) for the total cost is in: (A) [0, 10), (B) [10, 100), (C) [100, 1000), (D) [1000, 10000), (E) [10000, ∞). [§6.4]

Big concept: Var(aX + b) = a^2Var(X) — a shift adds nothing to variance, a scale multiplies it by the square.
Build the cost: 40 vaccines at $70 is a fixed $2800, and each of the X sick employees costs $1000, so C = 2800 + 1000X with X ~ B(40, 0.3) from A9 (Var X = 8.4).
Var(C) = 1000^2 Var(X) = 1,000,000(8.4) = **8,400,000**, in [10000, ∞). **(E).**
Why the $2800 vanishes: variance measures spread about the mean, and adding a constant moves the whole distribution without stretching it. Only the multiplier matters, and it enters squared because variance is in squared units.
Useful companions: E(C) = 2800 + 1000(12) = $14,800 (linearity, where the constant *does* count), and SD(C) = 1000(2.90) = $2898 — the SD scales by |a|, which is why SD is the figure you quote in dollars.

A11. Earthquakes follow a Poisson process at 4 per week. P(exactly 4 in the next 2 weeks) = (A) $e^{-2}2^4/4!$, (B) $e^{-4}4^4/4!$, (C) $e^{-8}8^4/4!$, (D) $e^{-2/4}(2/4)^4/4!$, (E) $e^{-4}4^2/2!$. [§9.4]

Big concept: in a Poisson process the count over an interval is Poisson with mean λt — you must scale the rate to the window.
The rate is λ = 4 per week and the window is t = 2 weeks, so N ~ Poisson(λt = 8), and
$P(N = 4) = \frac{e^{-8}8^4}{4!} \approx 0.0573$. **(C).**
How the distractors are built: (B) uses the one-week mean 4 (forgot to scale), (A) uses 2 (confused the time with the mean), (D) divides instead of multiplying, (E) mismatches the exponent and the factorial. The general pmf is $P(N = k) = e^{-\lambda t}(\lambda t)^k/k!$, and the same (λt) must appear in all three places.
Related facts worth carrying: counts on disjoint intervals are independent, E(N) = Var(N) = λt = 8, and the *waiting time* between earthquakes is Exponential(4) with mean 1/4 week.

A12. Brad passes 200 runners; each yells with probability 0.01. Using the Poisson approximation, P(yelled at more than twice) is in: (A) (0, 0.25], (B) (0.25, 0.40], (C) (0.40, 0.50], (D) (0.50, 0.70], (E) (0.70, 1]. [§9.2]

Big concept: Poisson approximation to the Binomial (n large, p small, $\lambda = np$), plus "more than" means take the complement.

Here n = 200 and p = 0.01 are exactly the regime where B(n, p) ≈ Poisson(np), with $\lambda = 200(0.01) = 2$.

$P(X > 2) = 1 - [P(0) + P(1) + P(2)] = 1 - e^{-2}(1 + 2 + \frac{4}{2}) = 1 - 5e^{-2} \approx 1 - 0.6767 =$ **0.3233**, in (0.25, 0.40].

(B).

Two places to be careful: (i) "more than twice" excludes 2, so you subtract k = 0, 1, 2 — subtracting only 0 and 1 gives the wrong 0.594; (ii) $P(2) = e^{-2}2^2/2! = 2e^{-2}$, so the bracket is 1 + 2 + 2 = 5, not 1 + 2 + 4.

Why the approximation is legitimate: the exact Binomial answer is 0.3233 as well — agreement to four digits, because np stays moderate while p is tiny. Use Poisson for rare events in many trials; use the Normal approximation instead when np and n(1−p) are both large.

A13. $p(x, y) = (x^2 + y^2)/140$ for x = 1, 2, 3 and y = −x, ..., x. Are X and Y independent? (A) yes, (B) no, (C) need more information. [§12.1]

Big concept: X and Y are independent only if $p(x, y) = p_X(x)p_Y(y)$ for *every* pair — which requires a rectangular support *and* a factoring formula.

Two independent reasons this fails:

1. *Support test (the fast one).* The range of y depends on x: y runs from −x to x. So p(1, 3) = 0 even though $p_X(1) > 0$ and $p_Y(3) > 0$, and the product rule already breaks. Whenever the admissible y values move with x, independence is impossible — no arithmetic needed.

2. *Factoring test.* Even on the support, $x^2 + y^2$ is a sum and cannot be written as g(x)h(y). A joint pmf of independent variables must split into an x-part times a y-part.

(B) no.

Caution about the converse: a rectangular support plus a factoring formula does imply independence, but a symmetric-looking pmf does not imply anything about covariance — here Cov(X, Y) = 0 by the ±y symmetry even though X and Y are dependent. Uncorrelated ≠ independent.

A14. 3 sophomores, 4 juniors, 5 seniors; 3 chosen at random; X = sophomores, Y = juniors. P(X = 1, Y = 1) = (A) $3 \cdot 4/\binom{12}{3}$, (B) $3 \cdot 4 \cdot 5/\binom{12}{3}$, (C) $(3 + 4)/\binom{12}{3}$, (D) (3/12)(4/12), (E) 3/12 + 4/12. [§11.2]

Big concept: sampling without replacement → multivariate hypergeometric — count favorable combinations over total combinations.

Three students are drawn from 3 + 4 + 5 = 12 without replacement, so all $\binom{12}{3} = 220$ triples are equally likely. X = 1 sophomore and Y = 1 junior forces the third pick to be one of the 5 seniors, so the favorable count is

$\binom{3}{1}\binom{4}{1}\binom{5}{1} = 3 \cdot 4 \cdot 5 = 60$,

giving P(X = 1, Y = 1) = 60/220 = 3/11 ≈ 0.2727. That is choice **(B)**, $3 \cdot 4 \cdot 5 \,/\, \binom{12}{3}$.

Why the other forms are wrong: (A) forgets the senior must be chosen too; (D) multiplies marginal proportions as if the draws were independent with replacement; (E) adds probabilities for events that are not alternatives. Order does not matter here because both numerator and denominator count unordered sets — keep them consistent.

Generalization to memorize: $P(k_1, k_2, k_3) = \binom{3}{k_1}\binom{4}{k_2}\binom{5}{k_3} \,/\, \binom{12}{3}$, the hypergeometric analogue of the multinomial (which is the with-replacement version, see A23).

A15. X, Y independent with MGFs M_X, M_Y. The MGF of $3X - 2Y$ is: (A) $M_X(t)^3 M_Y(t)^{-2}$, (B) $M_X(t)^3 + M_Y(t)^{-2}$, (C) $M_X(3t)M_Y(-2t)$, (D) $M_X(3t) + M_Y(-2t)$, (E) $3M_X(t) - 2M_Y(t)$. [§7.4]

Big concept: MGFs multiply over independent sums, and scaling moves *inside* the argument: $M_{aX}(t) = M_X(at)$.

By definition $M_Z(t) = E(e^{tZ})$. With $Z = 3X - 2Y$ and X, Y independent,

$M_Z(t) = E(e^{t(3X - 2Y)}) = E(e^{(3t)X})E(e^{(-2t)Y}) = M_X(3t)\, M_Y(-2t)$. **(C).**

Independence is what licenses splitting the expectation of the product into a product of expectations; the coefficients 3 and -2 then simply rescale t.

Why the distractors are tempting: (A) raises MGFs to powers — that is the rule for a sum of n *iid* copies, $M_{X_1+\cdots+X_n}(t) = M_X(t)^n$, a different situation. (B) and (D) add MGFs, which never corresponds to any operation on random variables. (E) confuses the MGF with expectation's linearity.

Why MGFs matter: they identify distributions uniquely, so this product rule is how you prove facts like "sum of independent Poissons is Poisson" or "sum of independent Normals is Normal". Also $M'(0) = E(X)$ and $M''(0) = E(X^2)$.

A16. Cov(4X − Y, X + Y) equals: (A) 4Var X + 3Cov + Var Y, (B) 16Var X + 3Cov − Var Y, (C) 4Var X + 3Cov − Var Y, (D) 16Var X + 4Cov + Var Y, (E) none. [§12.4]

Big concept: covariance is bilinear — expand it like a product, using Cov(X, X) = Var(X) and Cov(X, Y) = Cov(Y, X).

$\mathrm{Cov}(4X - Y, X + Y) = 4\mathrm{Cov}(X, X) + 4\mathrm{Cov}(X, Y) - \mathrm{Cov}(Y, X) - \mathrm{Cov}(Y, Y)$

$= 4\mathrm{Var}(X) + 4\mathrm{Cov}(X, Y) - \mathrm{Cov}(X, Y) - \mathrm{Var}(Y) =$ **$4\mathrm{Var}(X) + 3\mathrm{Cov}(X, Y) - \mathrm{Var}(Y)$. (C).**

The decisive detail: coefficients come out to the *first* power, so the 4 stays a 4 — distractors (B) and (D) use 16, importing the squaring rule from variance. Squaring is correct for $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ (see A10) because variance is covariance of a variable with itself, where the coefficient appears twice.

General template: $\mathrm{Cov}(aX + bY, cX + dY) = ac\mathrm{Var}(X) + (ad + bc)\mathrm{Cov}(X, Y) + bd\mathrm{Var}(Y)$. Setting $b = d = 1$, $a = 4$, $c = 1$ reproduces the answer. Also remember Cov(X, constant) = 0, so additive constants never matter.

A17. X ~ Poisson(1), Y ~ Poisson(2), independent. P(max(X, Y) ≤ 1) is in: (A) [0, 0.5), (B) [0.5, 0.65), (C) [0.65, 0.8), (D) [0.8, 0.9), (E) [0.9, 1). [§12.3]

Big concept: translate max/min events into statements about *all* the variables — max(X, Y) ≤ a means X ≤ a *and* Y ≤ a — then use independence to multiply the cdfs.

$P(\max(X, Y) \le 1) = P(X \le 1, Y \le 1) = P(X \le 1)P(Y \le 1)$.

For X ~ Poisson(1): $P(X \le 1) = e^{-1}(1 + 1) = 2e^{-1} \approx 0.7358$.

For Y ~ Poisson(2): $P(Y \le 1) = e^{-2}(1 + 2) = 3e^{-2} \approx 0.4060$.

Product: $6e^{-3} \approx$ **0.2987**, in [0, 0.5). **(A).**

The companion identity: min(X, Y) > a means both exceed a, so $P(\min > a) = P(X > a)P(Y > a)$ — use the max rule for upper bounds and the min rule for lower bounds, and complement whichever you need.

Sanity check: the answer must be smaller than either factor alone, and $0.2987 < 0.406$ ✓. Requiring two things at once can only shrink the probability.

A18. In proving the weak law of large numbers, Chebyshev bounds $P(|(X_1 + \cdots + X_n)/n - \mu| \geq \varepsilon)$ by: (A) $n\sigma^2/\varepsilon^2$, (B) $n\sigma^2/\varepsilon$, (C) $\sigma^2/(\varepsilon^2 n)$, (D) $\sigma^2/(\varepsilon n^2)$, (E) $\sigma^2/(\varepsilon^2 n^2)$. [§14.5]

Big concept: Chebyshev's inequality applied to the sample mean, with $\mathrm{Var}(\bar{X}) = \sigma^2/n$ — the engine of the weak law of large numbers.

Chebyshev says $P(|Z - E\,Z| \geq \varepsilon) \leq \mathrm{Var}(Z)/\varepsilon^2$. Take $Z = \bar{X} = (X_1 + \cdots + X_n)/n$, which has $E(\bar{X}) = \mu$ and, by independence plus the a^2 scaling rule,

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

Therefore $P(|\bar{X} - \mu| \geq \varepsilon) \leq \boldsymbol{\sigma^2/(n\varepsilon^2)}$. **(C).**

What this proves: the bound $\to 0$ as $n \to \infty$ for any fixed $\varepsilon > 0$, which is exactly the statement that $\bar{X}$ converges to μ in probability — the WLLN. Averaging n observations divides the variance by n, so the SD shrinks like $1/\sqrt{n}$.

Sign check on the distractors: any answer with n in the numerator (A, B) says averaging *more* data makes you *less* accurate, which is backwards; (D) and (E) misplace the squares — ε is always squared and n appears to the first power.

A19. Toothpick at $p = 1/4$; the bar is broken at a uniform U; Amy gets the piece with the toothpick, length g(U). $P(g(U) \geq 1/2)$ is in: (A) (0, 0.27], (B) (0.27, 0.45], (C) (0.45, 0.55], (D) (0.55, 0.65], (E) (0.65, 1]. [§10.1]

Big concept: for a continuous Uniform, probability = length of the favorable set — so translate the geometric condition into an interval for U.

The break point is $U \sim \mathrm{Uniform}(0, 1)$ and the toothpick sits at 1/4. Amy's piece is whichever side holds the toothpick:

- If $U > 1/4$, the toothpick is in the left piece, so $g(U) = U$.
- If $U < 1/4$, the toothpick is in the right piece, so $g(U) = 1 - U$.

Now impose $g(U) \geq 1/2$ in each case. When $U < 1/4$, $1 - U > 3/4 \geq 1/2$ automatically — the whole interval works, contributing 1/4. When $U > 1/4$, we need $U \geq 1/2$, contributing length 1/2. The two sets are disjoint, so $P(g(U) \geq 1/2) = 1/4 + 1/2 = \boldsymbol{3/4}$, in (0.65, 1]. **(E).**

The transferable move: a piecewise-defined function of a uniform is handled by splitting the sample space at the breakpoints, solving the inequality separately on each piece, and adding the lengths — never by plugging into a formula. The answer is large because a break on the short side hands Amy the long piece.

A20. Four athletes choose UNC/Duke/NCS independently with probabilities (0.5, 0.3, 0.2), (0.8, 0.1, 0.1), (0.3, 0.4, 0.3), (0.9, 0.05, 0.05). E(number of schools with no athlete) is in: (A) (0, 0.6], (B) (0.6, 1.2], (C) (1.2, 1.8], (D) (1.8, 2.4], (E) (2.4, 4]. [§11.4]

Big concept: expectation of a count = sum of indicator probabilities (linearity of expectation), which holds even when the indicators are dependent.

Let N = number of schools with no athlete, and write $N = I_{\mathrm{UNC}} + I_{\mathrm{Duke}} + I_{\mathrm{NCS}}$, where each indicator is 1 if that school gets nobody. Then $E(N) = P(\text{no one picks UNC}) + P(\text{no one picks Duke}) + P(\text{no one picks NCS})$, and each term is a product over the four independent athletes:

- UNC empty: $(0.5)(0.2)(0.7)(0.1) = 0.007$
- Duke empty: $(0.7)(0.9)(0.6)(0.95) = 0.3591$
- NCS empty: $(0.8)(0.9)(0.7)(0.95) = 0.4788$

$E(N) = 0.007 + 0.3591 + 0.4788 = \boldsymbol{0.8449}$, in (0.6, 1.2]. **(B).**

Why this is the whole trick: N's own distribution is messy (the three indicators are clearly dependent — if two schools are empty, everyone went to the third), but linearity of expectation needs no independence whatsoever. Independence is used only *inside* each term, to multiply the four athletes' probabilities.

Sanity check: UNC is nearly impossible to leave empty (every athlete has a decent chance of picking it), while Duke and NCS are each roughly coin-flips, so a total just under 1 is right.

A21. X uniform on [−0.5, 0.5], $Y = X^3$. $f_Y(0.05)$ is in: (A) (0, 1.5], (B) (1.5, 3], (C) (3, 4.5], (D) (4.5, 6], (E) (6, ∞). [§5.5]

Big concept: density of a transformed variable — for monotone g, $f_Y(y) = f_X(g^{-1}(y))\ |d\ g^{-1}(y)/dy|$ (change of variables).

X ~ Uniform(−0.5, 0.5) has $f_X(x) = 1$ on that interval. $Y = X^3$ is strictly increasing, so invert: $x = y^{1/3}$, and $\frac{dx}{dy} = \frac{1}{3}y^{-2/3}$.

Hence $f_Y(y) = 1 \cdot \frac{1}{3}y^{-2/3}$ for y in (−0.125, 0.125). At y = 0.05: $0.05^{2/3} \approx 0.1357$, so $f_Y(0.05) \approx 1/(3 \cdot 0.1357) \approx$ **2.456**, in (1.5, 3]. **(B).**

The cdf route (safer if you forget the formula): $F_Y(y) = P(X^3 \le y) = P(X \le y^{1/3}) = y^{1/3} + 0.5$ on the support; differentiate to get the same $(1/3)y^{-2/3}$. Deriving the density from the cdf always works and automatically handles the direction of the inequality — for a *decreasing* g the inequality flips, which is why the formula carries an absolute value.

Reality check: a density may exceed 1 (it is not a probability), and f_Y blows up near 0 because cubing squeezes a wide band of x values into a narrow band of y values near the origin.

A22. X questions with P(0, 1, 2) = 0.2, 0.4, 0.4; each answered inadequately with probability 1/5; Y = inadequate answers. P(XY ≤ 1) is in: (A) (0, 0.1], (B) (0.1, 0.2], (C) (0.2, 0.3], (D) (0.3, 0.6], (E) (0.6, 1.0]. [§11.2]

Big concept: a compound/conditional setup — condition on X, use the Binomial for Y given X, then recombine with the law of total probability. Unpack the event {XY ≤ 1} case by case.

Given X = x, Y ~ Binomial(x, 1/5). Take each value of X:

- X = 0 (prob 0.2): Y = 0, so XY = 0 ≤ 1 — always true, contributes 0.2.
- X = 1 (prob 0.4): XY = Y ∈ {0, 1}, which is ≤ 1 either way — always true, contributes 0.4.
- X = 2 (prob 0.4): XY = 2Y, which is ≤ 1 only if Y = 0, probability $(4/5)^2 = 0.64$ — contributes 0.4(0.64) = 0.256.

Total: 0.2 + 0.4 + 0.256 = **0.856**, in (0.6, 1.0]. **(E).**

Why case-by-case is unavoidable: XY ≤ 1 is not an event about X alone or Y alone, and X, Y are dependent (the number of trials for Y *is* X). Conditioning on X turns a dependent problem into three easy independent ones — the standard play whenever one variable sets the parameter of another.

Related computation worth practicing: E(Y) = E(E(Y | X)) = E(X/5) = (0 + 0.4 + 0.8)/5 = 0.24, using the tower property on the same decomposition.

A23. Five customers; preferences 0.1 shoyu, 0.1 shio, 0.5 miso, 0.3 tonkotsu. P(2 miso, 2 tonkotsu, 1 shoyu) is in: (A) (0, 0.001], (B) (0.001, 0.0025], (C) (0.0025, 0.1], (D) (0.1, 0.1125], (E) (0.1125, 0.2]. [§11.5]

Big concept: the multinomial distribution — independent trials with more than two outcome categories.

With n = 5 customers, each independently choosing among four ramen types with probabilities (0.1, 0.1, 0.5, 0.3),

$P(2 \text{ miso, } 2 \text{ tonkotsu, } 1 \text{ shoyu, } 0 \text{ shio}) = \frac{5!}{2!\,2!\,1!\,0!}(0.5)^2(0.3)^2(0.1)^1(0.1)^0$.

The multinomial coefficient is $120/(2 \cdot 2 \cdot 1 \cdot 1) = 30$, and the probability part is (0.25)(0.09)(0.1) = 0.00225. Multiply: 30(0.00225) = **0.0675**, in (0.0025, 0.1]. **(C).**

What the coefficient is doing: 0.00225 is the probability of one specific ordering (say miso, miso, tonkotsu, tonkotsu, shoyu); there are 30 distinguishable orderings of that multiset, and they are disjoint, so you add 30 copies. Forgetting the coefficient gives 0.00225 and answer (B) — the most common error here.

Connections: with two categories this is exactly the Binomial, and each marginal is Binomial (e.g. the miso count is B(5, 0.5)). The counts are negatively correlated, since one customer choosing miso means one fewer available for tonkotsu. Contrast A14, the without-replacement (hypergeometric) version.

A24. 8 left socks randomly paired with 8 right socks. Cov(I_{red}, I_{blue}) is in: (A) (−0.0025, −0.001], (B) (−0.001, 0], (C) (0, 0.001], (D) (0.001, 0.0025], (E) (0.0025, 0.1]. [§12.4]

Big concept: covariance of two indicator variables, Cov(I_A, I_B) = P(A ∩ B) − P(A)P(B) — and matching without replacement creates small dependence.
Randomly pair 8 left socks with 8 right socks. By symmetry, the red left sock is equally likely to meet any of the 8 right socks, so P(red matched) = 1/8, and likewise for blue.
For both: condition on the red pairing happening (1/8); the blue left sock now faces 7 remaining right socks, so P(blue matched | red matched) = 1/7. Hence P(both) = (1/8)(1/7) = 1/56.
Cov = $\frac{1}{56} - \frac{1}{64} = \frac{8-7}{448} = \frac{1}{448} \approx$ **0.00223**, in (0.001, 0.0025]. **(D).**
Why the covariance is positive, not negative: learning that red found its partner removes one "wrong" right sock from blue's pool, nudging blue's chance from 1/8 up to 1/7. Sampling-without-replacement dependence can go either way — here the competing socks are eliminated, which helps.
Where this is used: the matching problem's total count M = $\sum I_i$ has E(M) = 8(1/8) = 1 and Var(M) = $\sum$Var(I_i) + $\sum_{i\neq j}$Cov(I_i, I_j) = 8(7/64) + 56(1/448) = 1 — the indicator-plus-covariance method for variances of counts.

A25. Laptop lifetime Exponential with mean 36 months. The median is in: (A) [20, 30), (B) [30, 34), (C) [34, 36), (D) [36, 38), (E) [38, ∞). [§10.3]

Big concept: Exponential quantiles — solve F(m) = 1/2 to get median = (ln 2)/λ = (mean)(ln 2), which is always below the mean.
Mean 36 months means λ = 1/36, with F(t) = 1 − $e^{-t/36}$. Set F(m) = 0.5:
$e^{-m/36} = 0.5 \Rightarrow m = 36 \ln 2 \approx 36(0.6931) \approx$ **24.95** months, in [20, 30). **(A).**
Why median < mean: the Exponential is right-skewed — a few very long-lived laptops drag the mean up above the typical value. The ratio is fixed: median = 0.693 × mean for *every* Exponential, so answers near 36 (choices C, D) are reading the mean as the median.
Other facts to keep handy: SD = mean = 36 (so the coefficient of variation is 1), the survival function is the clean P(T > t) = $e^{-t/36}$, and memorylessness gives P(T > 50 | T > 14) = P(T > 36) — a used laptop is statistically as good as new.

16.2 Part B: a second exam

B1. 60% of email is spam. The word "free" appears in 30% of spam and 2% of non-spam. P(spam | contains "free") is in: (A) (0, 0.5], (B) (0.5, 0.8], (C) (0.8, 0.9], (D) (0.9, 0.95], (E) (0.95, 1]. [§2.2]

0.6(0.3)/[0.6(0.3) + 0.4(0.02)] = 0.18/0.188 ≈ 0.9574. **(E).**

B2. E, F, G independent with probabilities 0.5, 0.4, 0.3. P(exactly two of them occur) is in: (A) (0, 0.2], (B) (0.2, 0.25], (C) (0.25, 0.3], (D) (0.3, 0.4], (E) (0.4, 1]. [§3.3]

0.5(0.4)(0.7) + 0.5(0.6)(0.3) + 0.5(0.4)(0.3) = 0.14 + 0.09 + 0.06 = 0.29. **(C).**

B3. A and B alternate rolling one fair die, A first. A wins on rolling a 6; B wins on rolling a 5 or 6. P(A wins) is in: (A) (0, 0.3], (B) (0.3, 0.35], (C) (0.35, 0.4], (D) (0.4, 0.5], (E) (0.5, 1]. [§3.5]

(1/6)/[1 − (5/6)(2/3)] = (1/6)/(4/9) = 3/8 = 0.375. **(C).**

B4. X has cdf 0 for x < 1, 0.2 on [1, 2), 0.5 on [2, 4), 0.9 on [4, 7), 1 for x ≥ 7. Var(X) is in: (A) (0, 1], (B) (1, 2], (C) (2, 3], (D) (3, 4], (E) (4, ∞). [§4.3, §6.4]

pmf 0.2, 0.3, 0.4, 0.1 at 1, 2, 4, 7. E(X) = 3.1; E(X^2) = 0.2 + 1.2 + 6.4 + 4.9 = 12.7; Var = 12.7 − 9.61 = 3.09. **(D).**

B5. $f(x) = c(1 - x^2)$ on $[-1, 1]$. $P(X > 1/2)$ is in: (A) (0, 0.1], (B) (0.1, 0.2], (C) (0.2, 0.3], (D) (0.3, 0.5], (E) (0.5, 1]. [§5.3]

$\int_{-1}^{1}(1 - x^2) = 4/3$, so $c = 3/4$. $P(X > 1/2) = (3/4)[(x - x^3/3)]_{1/2}^{1} = (3/4)(2/3 - 11/24) = (3/4)(5/24) = 0.15625$. **(B).**

B6. A player makes 70% of free throws, independently. In 12 attempts, P(at least 10 made) is in: (A) (0, 0.15], (B) (0.15, 0.22], (C) (0.22, 0.28], (D) (0.28, 0.35], (E) (0.35, 1]. [§8]

$\binom{12}{10}(0.7)^{10}(0.3)^2 + 12(0.7)^{11}(0.3) + (0.7)^{12} \approx 0.1678 + 0.0712 + 0.0138 = 0.2528$. **(C).**

B7. 500 people each have a certain allergy with probability 0.004, independently. Using the Poisson approximation, P(exactly 3 have it) is in: (A) (0, 0.1], (B) (0.1, 0.15], (C) (0.15, 0.2], (D) (0.2, 0.3], (E) (0.3, 1]. [§9.2]

$\lambda = 2$: $e^{-2}2^3/3! = (8/6)e^{-2} \approx 0.1804$. **(C).**

B8. Customers arrive as a Poisson process at 6 per hour. P(at most one customer in a 20-minute span) is in: (A) (0, 0.2], (B) (0.2, 0.3], (C) (0.3, 0.35], (D) (0.35, 0.45], (E) (0.45, 1]. [§9.4]

$\lambda t = 6(1/3) = 2$: $e^{-2}(1 + 2) = 3e^{-2} \approx 0.4060$. **(D).**

B9. A bulb's lifetime is Exponential with mean 500 hours. Given it has lasted 500 hours, P(it lasts beyond 1000 hours) = (A) e^{-2}, (B) e^{-1}, (C) $1 - e^{-1}$, (D) $e^{-1/2}$, (E) 1/2. [§10.4]

Memoryless: $P(X > 500 \text{ more}) = e^{-500/500}$. **(B).**

B10. $X \sim$ Uniform(0, 10). $P(|X - 5| > 3)$ and Var(X) are: (A) 0.4 and 8.33, (B) 0.6 and 8.33, (C) 0.4 and 33.33, (D) 0.3 and 8.33, (E) 0.4 and 25. [§10.1]

$P(X < 2 \text{ or } X > 8) = 4/10$; $\text{Var} = 100/12 \approx 8.33$. **(A).**

B11. $M_X(t) = 0.3 + 0.5e^t + 0.2e^{3t}$. Var(X) is in: (A) (0, 0.5], (B) (0.5, 1], (C) (1, 1.2], (D) (1.2, 2], (E) (2, ∞). [§7.3]

$X = 0, 1, 3$ with probabilities 0.3, 0.5, 0.2. $E(X) = 1.1$, $E(X^2) = 0.5 + 1.8 = 2.3$, $\text{Var} = 2.3 - 1.21 = 1.09$. **(C).**

B12. $M_X(t) = e^{5(e^t - 1)}$. $P(X = 0)$ and Var(X) are: (A) e^{-5} and 5, (B) e^{-5} and 25, (C) $5e^{-5}$ and 5, (D) e^{-5} and $\sqrt{5}$, (E) $1 - e^{-5}$ and 5. [§7.3, §9]

Poisson(5). **(A).**

B13. Independent trials with success probability 0.3. P(the first success is on trial 4) is in: (A) (0, 0.05], (B) (0.05, 0.09], (C) (0.09, 0.11], (D) (0.11, 0.2], (E) (0.2, 1]. [§11.1]

$(0.7)^3(0.3) = 0.1029$. **(C).**

B14. Ten fair dice are rolled. E(number of distinct faces that appear) is in: (A) (0, 3], (B) (3, 4], (C) (4, 4.8], (D) (4.8, 5.2], (E) (5.2, 6]. [§11.4]

Indicator for each face: $P(\text{face appears}) = 1 - (5/6)^{10} \approx 0.8385$; times 6 gives ≈ 5.031. **(D).**

B15. A fair die is rolled 6 times. P(exactly two 1s, one 2 and three 6s) is in: (A) (0, 0.0005], (B) (0.0005, 0.001], (C) (0.001, 0.002], (D) (0.002, 0.01], (E) (0.01, 1]. [§11.5]

$6!/(2!\,1!\,3!) \cdot (1/6)^6 = 60/46656 \approx 0.00129$. **(C).**

B16. Var(X) = 4, Var(Y) = 9, Cov(X, Y) = −2. Var(2X − Y) and ρ(X, Y) are: (A) 33 and −1/3, (B) 25 and −1/3, (C) 17 and −1/3, (D) 33 and −2/13, (E) 25 and −1/6. [§12.5–12.6]

$4(4) + 9 - 2(2)(-2) = 33$; $\rho = -2/(2 \cdot 3) = -1/3$. **(A).**

B17. Joint pmf: p(0, 0) = 0.1, p(0, 1) = 0.2, p(0, 2) = 0.1, p(1, 0) = 0.2, p(1, 1) = 0.2, p(1, 2) = 0.2. Then X and Y are: (A) independent with Cov = 0, (B) not independent but Cov = 0, (C) not independent, Cov > 0, (D) not independent, Cov < 0, (E) independent with Cov ≠ 0. [§12.1, §12.4]

Marginals: P(X = 1) = 0.6; P(Y = 0, 1, 2) = 0.3, 0.4, 0.3. E(XY) = 0.2 + 0.4 = 0.6 = E(X)E(Y) = 0.6(1.0), so Cov = 0. But p(0, 0) = 0.1 ≠ 0.4(0.3) = 0.12, so not independent. **(B).**

B18. X, Y independent Uniform(0, 1). P(max(X, Y) ≤ 0.8) is in: (A) (0, 0.5], (B) (0.5, 0.6], (C) (0.6, 0.7], (D) (0.7, 0.8], (E) (0.8, 1]. [§12.3]

$0.8^2 = 0.64$. **(C).**

B19. X ~ Exponential(1) and Y ~ Exponential(3) are independent. P(min(X, Y) > 0.5) is in: (A) (0, 0.1], (B) (0.1, 0.2], (C) (0.2, 0.3], (D) (0.3, 0.5], (E) (0.5, 1]. [§12.3]

$P(\min > 0.5) = P(X > 0.5)P(Y > 0.5) = e^{-0.5}e^{-1.5} = e^{-2} \approx 0.1353$ (the minimum is Exponential(4)). **(B).**

B20. X ~ Binomial(3, 0.5) and Y ~ Binomial(2, 0.5) are independent. P(X + Y = 2) is in: (A) (0, 0.2], (B) (0.2, 0.3], (C) (0.3, 0.35], (D) (0.35, 0.5], (E) (0.5, 1]. [§12.2]

Same p, so X + Y ~ Binomial(5, 0.5): $\text{inom}\{5\}\{2\}/2^5 = 10/32 = 0.3125$. **(C).**

B21. X ~ N(10, 4). P(X > 13) is in: (A) (0, 0.05], (B) (0.05, 0.1], (C) (0.1, 0.2], (D) (0.2, 0.3], (E) (0.3, 1]. [§13.1]

$\sigma = 2$: $1 - \Phi(1.5) = 0.0668$. **(B).**

B22. A request goes to server A with probability 0.4 (errors Poisson(2)) or server B with probability 0.6 (errors Poisson(1)). P(no errors) is in: (A) (0, 0.2], (B) (0.2, 0.25], (C) (0.25, 0.3], (D) (0.3, 0.4], (E) (0.4, 1]. [§2.1, §9]

$0.4e^{-2} + 0.6e^{-1} \approx 0.0541 + 0.2207 = 0.2749$. **(C).**

B23. An urn has 5 red, 4 blue and 3 green balls; 4 are drawn without replacement. P(2 red, 1 blue, 1 green) is in: (A) (0, 0.1], (B) (0.1, 0.2], (C) (0.2, 0.3], (D) (0.3, 0.4], (E) (0.4, 1]. [§11.2]

$\binom{5}{2} \cdot 4 \cdot 3/\binom{12}{4} = 120/495 \approx 0.2424$. **(C).**

B24. X ~ Poisson(1.5) and Y ~ Poisson(2.5) are independent. P(X + Y ≤ 2) is in: (A) (0, 0.15], (B) (0.15, 0.2], (C) (0.2, 0.25], (D) (0.25, 0.35], (E) (0.35, 1]. [§12.2]

$X + Y \sim \text{Poisson}(4)$: $e^{-4}(1 + 4 + 8) = 13e^{-4} \approx 0.2381$. **(C).**

B25. A machine's number of cycles until failure is Geometric with p = 0.05. P(it survives more than 20 cycles), and P(it survives 20 more | it has survived 10), are: (A) 0.3585 and 0.3585, (B) 0.3585 and 0.1216, (C) 0.6415 and 0.3585, (D) 0.3585 and 0.5987, (E) 0.95 and 0.3585. [§11.1]

$P(X > 20) = 0.95^{20} \approx 0.3585$, and by memorylessness the conditional probability is the same. **(A).**

B26. X has mean 20 and variance 9. Chebyshev's upper bound on $P(|X - 20| \geq 6)$ is: (A) 1/4, (B) 1/2, (C) 3/6, (D) 9/6, (E) 1/36. [§14.4]

9/36. **(A).**

B27. A fair coin is flipped 400 times. Using the CLT with continuity correction, P(190 ≤ heads ≤ 210) is in: (A) (0, 0.5], (B) (0.5, 0.6], (C) (0.6, 0.68], (D) (0.68, 0.75], (E) (0.75, 1]. [§14.2]

Mean 200, SD 10: $\Phi(1.05) - \Phi(-1.05) = 2(0.8531) - 1 \approx 0.7063$. **(D).**

B28. A fair die is rolled 180 times; X = number of sixes. Using the Normal approximation with continuity correction, P(X ≥ 40) is in: (A) (0, 0.02], (B) (0.02, 0.05], (C) (0.05, 0.1], (D) (0.1, 0.2], (E) (0.2, 1]. [§14.2]

$X \sim \text{Bin}(180, 1/6)$: $\mu = 30$, $\sigma^2 = 25$, $\sigma = 5$. $P(W > 39.5) = 1 - \Phi(1.9) \approx 0.0287$ (exact Binomial 0.0321). **(B).**

17. One-page summary

Read this the night before and again right before the exam. The distribution table is the single most useful thing to have memorized.

17.1 The distributions

Name	Story	pmf / pdf	Mean	Variance	MGF
Bernoulli(p)	one yes/no trial	$P(1) = p$, $P(0) = 1 - p$	p	$p(1-p)$	$1 - p + pe^t$
Binomial(n, p)	successes in n independent trials	$\binom{n}{k}p^k(1-p)^{n-k}$, $k = 0..n$	np	$np(1-p)$	$(1 - p + pe^t)^n$
Poisson(λ)	rare events; Binomial with $\lambda = np$; events in a window of a rate-λ process	$e^{-\lambda}\lambda^k/k!$, $k = 0, 1, \ldots$	λ	λ	$e^{\lambda(e^t - 1)}$
Geometric(p)	trial number of the first success	$(1-p)^{k-1}p$, $k = 1, 2, \ldots$; $P(X > k) = (1-p)^k$	$1/p$	$(1-p)/p^2$	$pe^t/(1 - (1-p)e^t)$
Uniform(a, b)	random point in (a, b)	$1/(b-a)$ on (a, b); $F(x) = (x-a)/(b-a)$	$(a+b)/2$	$(b-a)^2/12$	$(e^{tb} - e^{ta})/(t(b-a))$
Exponential(λ)	waiting time at rate λ; memoryless	$\lambda e^{-\lambda x}$, $x > 0$; $P(X > x) = e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$	$\lambda/(\lambda - t)$, $t < \lambda$
Normal(μ, σ^2)	sums of many small effects; CLT limit	$e^{-(x-\mu)^2/2\sigma^2}/(\sigma\sqrt{(2\pi)})$	μ	σ^2	$e^{\mu t + \sigma^2 t^2/2}$
Multinomial(n; $p_{1..r}$)	n trials, r categories	$n!/(n_1! \cdot \cdots \cdot n_r!)\ \Pi\ p_i^{n_i}$	$E(X_i) = np_i$	$np_i(1-p_i)$; $\mathrm{Cov}(X_i, X_j) = -np_ip_j$	—

Sums of independent variables: Bin(n, p) + Bin(m, p) = Bin(n + m, p); Pois(λ) + Pois(μ) = Pois($\lambda + \mu$); $N(\mu_1, {\sigma_1}^2) + N(\mu_2, {\sigma_2}^2) = N(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2)$. Minimum of Exp($\lambda_1$) and Exp($\lambda_2$) is Exp($\lambda_1 + \lambda_2$). Poisson process: $N(t) \sim$ Pois(λt); time to the first event $\sim$ Exp(λ).

17.2 The formulas

CONDITIONING. $P(E|F) = P(E \cap F)/P(F)$. Chain: $P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)$. Total probability over a partition: $P(E) = \Sigma\ P(E|F_i)P(F_i)$. Bayes: $P(F_j|E) = P(E|F_j)P(F_j)/\Sigma_i P(E|F_i)P(F_i)$. Mammogram 9%; HIV low-risk 50%; negative test: use the negative-result rates.

INDEPENDENCE. $P(E \cap F) = P(E)P(F) \Leftrightarrow P(E|F) = P(E)$. Complements inherit it. Disjoint $\neq$ independent. n events: every sub-collection multiplies; events from disjoint groups are independent. Unions: go through the complement (De Morgan), then multiply. First-step: P(A before B) $= p_A/(p_A + p_B)$; alternating turns $p_A/(1 - (1-p_A)(1-p_B))$; "who fails first" = geometric series $r^{n-1} \cdot$ (decisive).

PMF, CDF, PDF. $p(a) = P(X = a)$, $\Sigma p = 1$. $F(x) = P(X \le x)$: staircase for discrete (jump location = value, jump size = probability), continuous for continuous. $P(a < X \le b) = F(b) - F(a)$. Continuous: $P(X \in B) = \int_B f$, $\int f = 1$, $F' = f$, $P(X = a) = 0$. Find c from $\int = 1$. $Y = g(X)$: $F_Y(y) = P(g(X) \le y)$, then differentiate.

Expectation. $E(X) = \Sigma\, x\ p(x)$ or $\int x\ f$. $E\ g(X) = \Sigma\ g(x)p(x)$ or $\int g\ f$ (plug-in). $E(I_A) = P(A)$. $E(aX + b) = aE(X) + b$. $E(X + Y) = E(X) + E(Y)$ always. Indicators: $E(\text{count}) = \Sigma\ P(A_i)$. Independent: $E(XY) = E(X)E(Y)$, $E\ g(X)h(Y) = E\ g(X)\ E\ h(Y)$.

Variance. $\text{Var}(X) = E(X-\mu)^2 = E(X^2) - (EX)^2 \geq 0$. $\text{SD} = \sqrt{\text{Var}}$. $\text{Var}(aX + b) = a^2\text{Var}(X)$. $\text{Var}(X \pm Y) = \text{Var}\ X + \text{Var}\ Y \pm 2\text{Cov}(X, Y)$. $\text{Var}(aX + bY) = a^2\text{Var}\ X + b^2\text{Var}\ Y + 2ab\ \text{Cov}$. Independent/uncorrelated: variances add. $E(X^2) = \text{Var} + (EX)^2$. $E(X' - X'')^2 = 2\text{Var}(X)$. Die: 35/12.

Covariance. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$; $\text{Cov}(X, X) = \text{Var}\ X$; bilinear; constants vanish. Independent $\Rightarrow$ Cov = 0, not conversely. $\rho = \text{Cov}/(\sigma_X\sigma_Y) \in [-1, 1]$. Indicators: $\text{Cov} = P(A\cap B) - P(A)P(B)$. Multinomial: $-np_ip_j$.

MGF. $M(t) = E\ e^{tX}$; $M(0) = 1$; $M'(0) = EX$; $M''(0) = EX^2$. Determines the distribution: read "Σ prob $\cdot$ $e^{t\,\cdot\,\text{value}}$" as a pmf. $M_{aX+b}(t) = e^{bt}M_X(at)$. Independent: $M_{X+Y} = M_XM_Y$, $M_{aX+bY}(t) = M_X(at)M_Y(bt)$. $E\ e^{X+Y} = M_X(1)M_Y(1)$.

Joint. Joint cdf $F(a, b) = P(X \leq a, Y \leq b)$; $P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F(a, b)$. $p(x, y) = P(X = x, Y = y)$; marginals = row/column sums; $p_{X|Y} = p/p_Y$; $p(x, y) = p_X(x)p_{Y|X}(y|x)$. Independent $\Leftrightarrow$ $p = p_Xp_Y$ for all cells $\Leftrightarrow$ product form on a rectangle. Max/min: $P(\max \leq a) = \Pi P(X_i \leq a)$, $P(\min > a) = \Pi P(X_i > a)$. Continuous: f(x, y), marginals by integrating.

Normal. $(X - \mu)/\sigma \sim N(0, 1)$. $P(a < X < b) = \Phi((b-\mu)/\sigma) - \Phi((a-\mu)/\sigma)$; $\Phi(-z) = 1 - \Phi(z)$. $\Phi(1) = .8413$, $\Phi(1.645) = .95$, $\Phi(1.96) = .975$, $\Phi(2) = .9772$, $\Phi(3) = .9987$. 68–95–99.7. $E|Z| = \sqrt{(2/\pi)} = .7979$; $E\ Z^4 = 3$. Linear combos of independent Normals are Normal.

Inequalities and limits. Markov ($X \geq 0$): $P(X \geq a) \leq EX/a$. Chebyshev: $P(|X - \mu| \geq a) \leq \sigma^2/a^2$. $\bar{X}$: mean μ, variance σ^2/n. WLLN: $P(|\bar{X} - \mu| \geq \varepsilon) \leq \sigma^2/(n\varepsilon^2) \to 0$. CLT: $(S_n - n\mu)/(\sigma\sqrt{n}) \to N(0, 1)$; $S_n \approx N(n\mu, n\sigma^2)$, $\bar{X} \approx N(\mu, \sigma^2/n)$. Binomial $\approx N(np, np(1-p))$ with ± 0.5 continuity correction.

Numbers to have ready. $e^{-1} = .3679$, $e^{-2} = .1353$, $e^{-3} = .0498$, $e^{-4} = .0183$, $e^{-5} = .0067$, $\ln 2 = .6931$, $\sqrt{(2\pi)} = 2.5066$, $\sqrt{(2/\pi)} = .7979$. Pois(λ): $P(0) = e^{-\lambda}$, $P(\geq 1) = 1 - e^{-\lambda}$, $P(>2) = 1 - e^{-\lambda}(1 + \lambda + \lambda^2/2)$. Exponential median $= (\ln 2)/\lambda$. Geometric series from k: $x^k/(1 - x)$.

Exam. Thursday Nov 12, in class, closed book, calculator with e^x, n!, nCr. About 12 multiple-choice questions: compute to four decimals and pick the interval (check endpoints), or pick the formula (test at easy values). Write the model and its parameters before computing.

Built November 12, 2026 from the STOR 435 course materials (Fall 2026, UNC): slide decks 5–8 with in-class notes, homeworks 6–10, the posted Midterm 2 practice problems, the instructor's Lectures 5–8 study guide and the syllabus. Personal study material.

© 2026 Deven Jadhav. Not affiliated with or endorsed by UNC-Chapel Hill; course material © its respective owners. Willful large-scale copyright infringement may be a federal crime (17 U.S.C. §506).