Probability Theory Final Exam Cheat Sheet

December 16, 2025 — 7:10pm-8:40pm — 3 Questions, 1.5 Hours

Open Book Exam - Focus on Post-Midterm 2 Material

0. QUICK REFERENCE GUIDE

TERMINOLOGY TRAPS (CRITICAL!)

- **"Gaussian" = Normal!** $N(\mu, \sigma^2)$
- **"Gaussian vector" = MVN** (Multivariate Normal)
- **"Independent components"** $= \rho = 0$ (for MVN: independence!)
- **"Mean $\theta = 3$" (Exp)** $\Rightarrow \lambda = 1/3$ NOT 3!
- For MVN ONLY: $\rho = 0 \Leftrightarrow$ independent
- $\psi(t)$ = MGF (professor's notation)
- **"Jointly normal"** = Same as Gaussian vector/MVN
- **"Rate parameter"** vs **"Scale parameter"**: Exp has λ (rate), mean $= 1/\lambda$
- **"Variance σ^2"** vs **"Standard deviation σ"**: $N(\mu, \sigma^2)$ uses variance!
- **"Proportion"** or **"probability parameter"** $\rightarrow$ Beta distribution
- **"Failures before success"** = Geometric (our convention)
- **"Trials until success"** = Geometric + 1 (alternate convention)

QUICK CALCULATION SHORTCUTS

- **Variance shortcut:** $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ (faster than definition!)
- **Sum of i.i.d.:** $E[\sum X_i] = n\mu$, $\mathrm{Var}(\sum X_i) = n\sigma^2$
- **Sample mean:** $E[\bar{X}] = \mu$, $\mathrm{Var}(\bar{X}) = \sigma^2/n$
- **Linear transform:** $E[aX + b] = aE[X] + b$, $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
- **Covariance of sum:** $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$
- **Independent:** $\mathrm{Cov}(X, Y) = 0$, so $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
- **Standardize:** $Z = (X - \mu)/\sigma \sim N(0, 1)$
- **De-standardize:** $X = \mu + \sigma Z$

VISUAL DECISION TREE (Start Here!)

```
START: What type of problem?
  |
  +--[Named Distribution?]
  |     +--"Gaussian/Normal" -> Sec 3.3 (single) or 4.5 (joint)
  |     +--"Exponential" -> Sec 3.4 [lambda=1/mean!]
  |     +--"Poisson" -> Sec 2.3
  |     +--"Binomial" -> Sec 2.2
  |     +--"Lognormal" -> Sec 7.3 (stock prices!)
  |     +--"Beta" -> Sec 3.6 (priors!)
  |
  +--[Multiple Variables?]
  |     +--"Joint PDF/PMF" -> Sec 4.1
  |     +--"Bivariate Normal/Gaussian vector" -> Sec 4.5
  |     +--"X+Y, sum" -> MGF (5.2) or direct
  |     +--"Max/Min" -> Order Stats (4.7)
  |     +--"X|Y=y" -> Conditional (4.2, 4.5)
  |
  +--[Approximation/Limits?]
  |     +--"Large n/approximate" -> CLT (6.1)
  |     +--"Sample mean" -> CLT (6.1)
  |     +--"n games/trials" -> CLT (6.1) Template B
  |
  +--[Bayesian?]
  |     +--"Prior/Posterior" -> Sec 7.2
  |     +--"Update belief" -> Bayes (1.3, 7.2)
  |     +--"Defective rate" -> Discrete Bayes (8.8)
  |     +--"Monty Hall" -> Template J
  |
  +--[Expectation?]
        +--"E[X|Y]" -> Cond. Expectation (7.1)
        +--"Total Expectation" -> E[X]=E[E[X|Y]]
```

Emergency Quick Reference: Problem Phrase → Section

- **"Gaussian"** → Normal! Sec 3.3
- **"Gaussian vector"** → MVN! Sec 4.5
- **"Independent components"** → $\rho = 0$ for MVN, Sec 4.5
- **"Large n"** / **"Approximate"** → CLT, Sec 6.1
- **"i.i.d."** → Independence, maybe CLT
- **"Prior/Posterior"** → Bayesian, Sec 7.2
- **"Update belief"** → Bayes' Theorem, Sec 1.3
- **"Conjugate"** → Beta-Binomial, Sec 7.2
- **"Stock price"** / **"$S_0 e^Z$"** → Lognormal, Sec 7.3
- **"Mean θ" (Exp)** → $\lambda = 1/\theta$! Sec 3.4
- **"Memoryless"** → Exponential, Sec 3.4
- **"Arrival/Counting process"** → Poisson, Sec 2.3
- **"Max/Min of n"** → Order Statistics, Sec 4.7
- **"$\psi(t)$"** → MGF! Sec 5.1
- **"Conditional distribution"** → Sec 4.2, 4.5
- **"$E[X|Y]$"** → Conditional Expectation, Sec 7.1
- **"Total winnings/games"** → CLT, Template B
- **"Monty Hall"** → Bayesian, Sec 8.1
- **"Defective rate"** → Bayesian, Sec 8.8

Top 20+ Critical Formulas

1. $\boxed{P(A|B) = \dfrac{P(A \cap B)}{P(B)}}$
2. $\boxed{P(H|E) = \dfrac{P(E|H)P(H)}{P(E)}}$ (Bayes)
3. $\boxed{P(A) = \sum P(A|B_i)P(B_i)}$ (Total Prob)
4. $\boxed{E[X] = \sum xP(X = x)}$ (Discrete)
5. $\boxed{E[X] = \int xf(x)dx}$ (Continuous)
6. $\boxed{\mathrm{Var}(X) = E[X^2] - (E[X])^2}$
7. $\boxed{\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]}$
8. $\boxed{\rho = \dfrac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}}$
9. $\boxed{P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}}$ (Binomial)
10. $\boxed{P(X = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}}$ (Poisson)

11. $\boxed{f(x) = \lambda e^{-\lambda x}, x > 0}$ (Exponential)

12. $\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$ (Normal)

13. $\boxed{Z = \frac{X - \mu}{\sigma}}$ (Standardization)

14. $\boxed{M(t) = E[e^{tX}]}$ (MGF)

15. $\boxed{E[X] = E[E[X|Y]]}$ (Total Expectation)

16. $\boxed{\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])}$

17. $\boxed{Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)}$ (CLT)

18. $\boxed{\text{CI} : \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$

19. $\boxed{\pi(\theta|x) \propto L(x|\theta)\pi(\theta)}$ (Bayes)

20. $\boxed{E[e^X] = e^{\mu+\sigma^2/2}}$ (Lognormal)

21. $\boxed{f_{UV}(u, v) = f_{XY}(x, y)|J|}$ (Jacobian)

Exponential Quick Reference

- PDF: $f(x) = \lambda e^{-\lambda x}$, CDF: $F(x) = 1 - e^{-\lambda x}$
- Mean: $E[X] = 1/\lambda$, Var: $1/\lambda^2$, SD: $1/\lambda$
- **! Mean $\theta = 3 \Rightarrow \lambda = 1/3$**
- Memoryless: $P(X > s + t|X > s) = P(X > t)$
- $P(X > x) = e^{-\lambda x}$

Common Mistakes Checklist

☐ Forgot continuity correction for discrete→normal
☐ Confused $P(A|B)$ with $P(B|A)$
☐ Didn't check independence before using formulas
☐ Wrong integration limits for marginals
☐ Forgot to normalize Bayesian posterior
☐ Used Binomial instead of Hypergeometric
☐ Forgot absolute value of Jacobian
☐ Assumed correlation implies causation
☐ Used λ instead of $1/\lambda$ for Exp mean
☐ Forgot that $\text{Var}(aX) = a^2\text{Var}(X)$ (not a)
☐ Used wrong variance formula: σ^2/n vs $\sigma^2 \cdot n$
☐ Forgot to square σ in variance formula
☐ Assumed uncorrelated means independent (only true for MVN!)
☐ Used $P(X = x) > 0$ for continuous (it's always 0!)
☐ Forgot that MGF of sum = product of MGFs (independent only!)
☐ Mixed up $\Phi(z)$ and $1 - \Phi(z)$ for tail probabilities
☐ Forgot that $\Phi(-z) = 1 - \Phi(z)$ (symmetry)
☐ Used $z_{0.05} = 1.96$ instead of $z_{0.025} = 1.96$ for 95% CI

Standard Normal Table Quick Values

- $\Phi(0) = 0.5$, $\Phi(1) = 0.8413$, $\Phi(2) = 0.9772$, $\Phi(3) = 0.9987$
- $\Phi(-z) = 1 - \Phi(z)$ (symmetry property)
- $z_{0.10} = 1.282$, $z_{0.05} = 1.645$, $z_{0.025} = 1.96$, $z_{0.01} = 2.326$, $z_{0.005} = 2.576$
- 68-95-99.7 rule: $P(|Z| < 1) = 0.68$, $P(|Z| < 2) = 0.95$, $P(|Z| < 3) = 0.997$

1. FUNDAMENTAL CONCEPTS

1.1 Probability Axioms

Terms: S = sample space (all outcomes), A, B = events (subsets of S), $P(A)$ = probability of A, A^c = complement (not A), $A \cap B$ = intersection (both), $A \cup B$ = union (either)

- **Definition:** A probability measure satisfies:
  1. Normalization: $P(S) = 1$
  2. Non-negativity: $P(A) \geq 0$
  3. Additivity: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$
- **Key Formulas:**
  - $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$
  - $\boxed{P(A^c) = 1 - P(A)}$
  - $\boxed{P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) -}$
- **When to Use:** Basic probability calculations, "or" problems
- **Solution Steps:**
  1. Identify sample space S and list all outcomes
  2. Define events A, B clearly
  3. Check if events are mutually exclusive ($A \cap B = \emptyset$?)
  4. If YES: $P(A \cup B) = P(A) + P(B)$
  5. If NO: Use inclusion-exclusion
  6. Count favorable outcomes or calculate probabilities
- **Worked Example 1:** Two dice rolled. Find $P(\text{sum} = 7)$.
  - Sample space: $|S| = 36$ equally likely outcomes
  - Favorable: $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$ = 6 outcomes
  - $P(\text{sum} = 7) = 6/36 = 1/6$
- **Worked Example 2:** $P(A) = 0.4$, $P(B) = 0.5$, $P(A \cap B) = 0.2$. Find $P(A \cup B)$.
  - $P(A \cup B) = 0.4 + 0.5 - 0.2 = 0.7$
  - Also: $P(A^c \cap B^c) = P((A \cup B)^c) = 1 - 0.7 = 0.3$
- **Worked Example 3:** Roll a die. $A = \{\text{even}\}$, $B = \{> 3\}$. Find $P(A \cup B)$.
  - $A = \{2, 4, 6\}$, $P(A) = 3/6 = 1/2$
  - $B = \{4, 5, 6\}$, $P(B) = 3/6 = 1/2$
  - $A \cap B = \{4, 6\}$, $P(A \cap B) = 2/6 = 1/3$
  - $P(A \cup B) = 1/2 + 1/2 - 1/3 = 2/3$
- **Common Pitfalls:**
  - Forgetting the intersection term when events overlap

- Assuming events are mutually exclusive without checking
- Not listing sample space carefully

- *Note: Equally likely:* $P(A) = |A|/|S|$. *DeMorgan:* $(A \cup B)^c = A^c \cap B^c$

1.2 Conditional Probability

Terms: $P(A|B)$ = probability of A given B occurred, A = target event, B = conditioning event (what we know happened)

- **Definition:** Probability of A given B occurred
- **Key Formulas:**
  - $\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) > 0}$
  - $\boxed{P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)}$ (Multiplication Rule)
  - $\boxed{P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)}$ (Chain Rule)
- **When to Use:** "Given that", "if we know", "conditional on", "among those who"
- **Solution Steps:**
  1. Identify: What is the **condition** B? What is the **target** A?
  2. Method 1 (Definition): Find $P(A \cap B)$ and $P(B)$ separately, divide
  3. Method 2 (Reduced sample space): Restrict to outcomes in B, count A within B
  4. Check: Does answer make sense? Is it between 0 and 1?
- **Worked Example 1:** Roll two dice. Find $P(\text{sum} < 8|\text{sum is odd})$.
  - $B = \{\text{sum odd}\}$: outcomes with sum 3,5,7,9,11 $\Rightarrow |B| = 18$
  - $A = \{\text{sum} < 8\}$: outcomes with sum 3,5,7 $\Rightarrow |A \cap B| = 12$
  - $P(A|B) = 12/18 = 2/3$
- **Worked Example 2:** Draw 2 cards without replacement. $P(\text{2nd is Ace}|\text{1st is Ace})$?
  - After drawing 1 Ace: 51 cards remain, 3 are Aces
  - $P(\text{2nd Ace}|\text{1st Ace}) = 3/51 = 1/17$
- **Worked Example 3:** $P(A) = 0.3$, $P(B) = 0.4$, $P(A|B) = 0.5$. Find $P(B|A)$.
  - First find $P(A \cap B) = P(B) \cdot P(A|B) = 0.4 \times 0.5 = 0.2$
  - Then $P(B|A) = P(A \cap B)/P(A) = 0.2/0.3 = 2/3$
- **Common Pitfalls:**
  - **! Confusing $P(A|B)$ with $P(B|A)$** – These are usually different!
  - Forgetting that conditioning changes the sample space
  - Using $P(A \cap B) = P(A)P(B)$ when events are NOT independent
- *Note:* $P(A|B) = P(A)$ *iff* A *and* B *are independent.* $P(A|B) + P(A^c|B) = 1$

1.3 Bayes' Theorem

Terms: H_i = hypothesis i, E = evidence (observed data), $P(H_i)$ = prior (belief before evidence), $P(E|H_i)$ = likelihood (prob of evidence given hypothesis), $P(H_i|E)$ = posterior (updated belief after evidence)

- **Definition:** Update probability given evidence (prior → posterior)
- **Key Formulas:**
  - $\boxed{P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)}}$ (Full Bayes)
  - $\boxed{P(H|E) = \frac{P(E|H)P(H)}{P(E)}}$ (Two hypothesis case)
  - $\boxed{P(E) = \sum_j P(E|H_j)P(H_j)}$ (Law of Total Probability)
- **Terminology:**
  - **Prior:** $P(H_i)$ = belief before evidence
  - **Likelihood:** $P(E|H_i)$ = prob of evidence if hypothesis true
  - **Posterior:** $P(H_i|E)$ = updated belief after evidence
- **When to Use:** "Update", "posterior", "given evidence", "test positive", "defective"
- **Solution Steps (Table Method):**
  1. List all hypotheses $H_1, H_2, \ldots$ in rows
  2. Column 1: Prior $P(H_i)$
  3. Column 2: Likelihood $P(E|H_i)$ – **prob of evidence GIVEN hypothesis**
  4. Column 3: Joint = $P(H_i) \times P(E|H_i)$
  5. Sum Column 3 to get $P(E)$ (denominator)
  6. Column 4: Posterior = Joint$/P(E)$
  7. Verify: Posteriors sum to 1
- **Worked Example:** Disease test. 1% have disease. Test: 95% true positive, 10% false positive. Person tests positive. $P(\text{disease}|+)$?
  - H_1: Disease, $P(H_1) = 0.01$, $P(+|H_1) = 0.95$, Joint = 0.0095
  - H_2: No disease, $P(H_2) = 0.99$, $P(+|H_2) = 0.10$, Joint = 0.099
  - $P(+) = 0.0095 + 0.099 = 0.1085$
  - $P(\text{disease}|+) = 0.0095/0.1085 = 0.0875 \approx 8.8\%$
- **Worked Example 2:** Two urns. Urn A: 3 red, 2 blue. Urn B: 1 red, 4 blue. Pick urn uniformly, draw red ball. $P(\text{Urn A}|\text{red})$?
  - $P(A) = P(B) = 1/2$ (prior)
  - $P(\text{red}|A) = 3/5$, $P(\text{red}|B) = 1/5$
  - $P(\text{red}) = (1/2)(3/5) + (1/2)(1/5) = 4/10 = 2/5$
  - $P(A|\text{red}) = \frac{(3/5)(1/2)}{2/5} = \frac{3/10}{2/5} = 3/4$
- **Common Pitfalls:**
  - **! Confusing likelihood $P(E|H)$ with posterior $P(H|E)$**

* Forgetting to normalize (posteriors must sum to 1)
* Using prior as likelihood

– *Note: Law of Total Probability:* $P(A) = \sum_i P(A|B_i)P(B_i)$ *where* B_i *partition sample space*

1.4 Independence

Terms: Independent = one event doesn't affect the other's probability, Pairwise independent = any two events in a set are independent, Mutually independent = all combinations of events are independent, $A \cap B$ = both A and B occur

– **Definition:** A and B independent if $P(A \cap B) = P(A)P(B)$

– **Equivalent Conditions (any one implies others):**
  * $P(A \cap B) = P(A)P(B)$
  * $P(A|B) = P(A)$ (knowing B doesn't change A)
  * $P(B|A) = P(B)$ (knowing A doesn't change B)

– **When to Use:** "Are events independent?", "does A affect B?", "with/without replacement"

– **Solution Steps to TEST Independence:**
  1. Calculate $P(A)$, $P(B)$ separately
  2. Calculate $P(A \cap B)$ directly
  3. Check: Is $P(A \cap B) = P(A) \cdot P(B)$?
  4. If YES: Independent. If NO: Dependent.

– **Worked Example 1:** Roll die. $A = \{\text{even}\}$, $B = \{1, 2, 3, 4\}$. Independent?
  * $P(A) = 3/6 = 1/2$, $P(B) = 4/6 = 2/3$
  * $A \cap B = \{2, 4\}$, so $P(A \cap B) = 2/6 = 1/3$
  * Check: $P(A)P(B) = (1/2)(2/3) = 1/3 = P(A \cap B)$ ✓
  * **Conclusion: Independent!**

– **Worked Example 2:** Roll die. $A = \{\text{even}\}$, $C = \{6\}$. Independent?
  * $P(A) = 1/2$, $P(C) = 1/6$
  * $A \cap C = \{6\}$, so $P(A \cap C) = 1/6$
  * Check: $P(A)P(C) = (1/2)(1/6) = 1/12 \neq 1/6$
  * **Conclusion: Dependent!**

– **Types of Independence:**
  * **Pairwise:** Each pair is independent
  * **Mutual:** $P(A \cap B \cap C) = P(A)P(B)P(C)$ AND all pairs
  * **! Pairwise independence does NOT imply mutual independence!**

– **Independence for Random Variables:**
  * X, Y independent iff $f(x, y) = f_X(x) \cdot f_Y(y)$ for all x, y
  * If independent: $E[XY] = E[X]E[Y]$, $\text{Cov}(X, Y) = 0$
  * **! $\text{Cov}(X, Y) = 0$ does NOT imply independence (except for Normal!)**

– **Common Pitfalls:**
  * Assuming independence without verification
  * Confusing "mutually exclusive" with "independent" (opposites!)
  * Assuming zero correlation implies independence

– *Note: Mutually exclusive ⇒ dependent (unless one has prob 0). If* $A \cap B = \emptyset$*, then* $P(A \cap B) = 0 \neq P(A)P(B)$ *unless* $P(A) = 0$ *or* $P(B) = 0$*.*

1.5 Counting Methods

Terms: n = total items, k = items chosen, $n!$ = factorial ($n \times (n-1) \times \cdots \times 1$), $P(n, k)$ = permutations (order matters), $\binom{n}{k}$ = combinations (order doesn't matter), Replacement = item can be chosen again

– **Decision Tree:** Does order matter? With/without replacement?
  * Order matters + with replacement: n^k
  * Order matters + without replacement: $P(n, k) = \frac{n!}{(n-k)!}$
  * Order doesn't matter + without replacement: $\binom{n}{k}$
  * Order doesn't matter + with replacement: $\binom{n+k-1}{k}$ (stars and bars)

– **Permutations (Order matters):**

$$\boxed{P(n, k) = \frac{n!}{(n-k)!} = n(n-1)(n-2)\cdots(n-k+1)}$$

– **Combinations (Order doesn't matter):**

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{P(n, k)}{k!}}$$

– **Multinomial (Partition into groups):**

$$\boxed{\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}}$$

– **Useful Identities:**
  * $\binom{n}{k} = \binom{n}{n-k}$ (symmetry)
  * $\binom{n}{0} = \binom{n}{n} = 1$
  * $\binom{n}{1} = n$
  * $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ (Pascal's triangle)

– **Worked Example 1:** How many ways to arrange letters in "MISSISSIPPI"?
  * 11 letters: M(1), I(4), S(4), P(2)
  * Answer: $\frac{11!}{1! \cdot 4! \cdot 4! \cdot 2!} = 34650$

– **Worked Example 2:** Committee of 5 from 10 people. How many ways?
  * Order doesn't matter, no replacement
  * Answer: $\binom{10}{5} = \frac{10!}{5!5!} = 252$

– **Worked Example 3:** 3-digit codes using digits 0-9 with repetition allowed?
  * Order matters, with replacement
  * Answer: $10^3 = 1000$

– **Worked Example 4:** Choose 3 from $\{A, B, C, D, E\}$ where order matters, no repetition?
  * Permutation: $P(5, 3) = 5 \times 4 \times 3 = 60$

– **Common Pitfalls:**
  * Confusing permutation vs combination (order matters?)
  * Forgetting "without replacement" constraint
  * Double counting in complex problems

– *Note: Probability:* $P = \frac{\textit{favorable}}{\textit{total}} = \frac{\textit{count ways for event}}{\textit{count all outcomes}}$

2. DISCRETE RANDOM VARIABLES

2.1 PMF and CDF

Terms: PMF = Probability Mass Function $p(x) = P(X = x)$, CDF = Cumulative Distribution Function $F(x) = P(X \leq x)$, $E[X]$ = expected value (mean), $\text{Var}(X)$ = variance, $\text{SD}(X)$ = standard deviation, LOTUS = Law of the Unconscious Statistician

- **PMF (Probability Mass Function):**
  - $p(x) = P(X = x)$ for each possible value x
  - Must satisfy: $p(x) \geq 0$ and $\sum_{\text{all } x} p(x) = 1$
- **CDF (Cumulative Distribution Function):**
  - $F(x) = P(X \leq x) = \sum_{k \leq x} p(k)$
  - Properties: Right-continuous, non-decreasing, $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$
- **Key Formulas:**
  - $\boxed{E[X] = \sum_x x \cdot P(X = x)}$
  - $\boxed{E[g(X)] = \sum_x g(x) \cdot P(X = x)}$ (LOTUS)
  - $\boxed{\text{Var}(X) = E[X^2] - (E[X])^2}$
  - $\boxed{\text{SD}(X) = \sqrt{\text{Var}(X)}}$
- **CDF to Probability Conversions:**
  - $P(X \leq a) = F(a)$
  - $P(X < a) = F(a^-) = \lim_{x \to a^-} F(x)$ (left limit)
  - $P(X > a) = 1 - F(a)$
  - $P(X \geq a) = 1 - F(a^-)$
  - $P(a < X \leq b) = F(b) - F(a)$
  - $P(X = a) = F(a) - F(a^-)$ (jump at a)
- **Worked Example:** X takes values $\{1, 2, 3\}$ with $P(X = k) = k/6$.
  - Verify: $1/6 + 2/6 + 3/6 = 1$ ✓
  - $E[X] = 1(1/6)+2(2/6)+3(3/6) = 1/6+4/6+9/6 = 14/6 = 7/3$
  - $E[X^2] = 1(1/6) + 4(2/6) + 9(3/6) = 1/6 + 8/6 + 27/6 = 36/6 = 6$
  - $\text{Var}(X) = 6 - (7/3)^2 = 6 - 49/9 = 54/9 - 49/9 = 5/9$
  - CDF: $F(1) = 1/6$, $F(2) = 1/6 + 2/6 = 1/2$, $F(3) = 1$
- **How to Construct PMF from Word Problem:**
  1. List all possible values X can take
  2. For each value, calculate $P(X = x)$
  3. Verify probabilities sum to 1
  4. Present as table or formula
- *Note: For discrete RV: $P(X = a) > 0$ possible. For continuous: $P(X = a) = 0$ always!*

2.2 Binomial Distribution (a.k.a. Binomial(n, p), "n trials")

Terms: n = number of trials, p = probability of success on each trial, $q = 1 - p$ = probability of failure, k = number of successes, Bernoulli trial = single trial with success/failure outcome, MGF = Moment Generating Function

- **SYNONYMS:** "n trials", "success/failure", "fixed number of trials", "with replacement"
- **Definition:** Number of successes in n independent Bernoulli trials with success prob p
- **PMF:** $\boxed{P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, ..., n}$
- **Parameters and Moments:**
  - Mean: $E[X] = np$
  - Variance: $\text{Var}(X) = np(1 - p) = npq$ where $q = 1 - p$
  - Mode: $\lfloor (n + 1)p \rfloor$ or $\lfloor (n + 1)p \rfloor - 1$
  - MGF: $M(t) = (1 - p + pe^t)^n = (q + pe^t)^n$
- **Conditions for Binomial (MUST ALL HOLD):**
  1. Fixed number of trials n
  2. Each trial: success or failure (binary)
  3. Constant probability p for each trial
  4. Trials are independent
- **Worked Example 1:** Flip fair coin 10 times. $P(\text{exactly 6 heads})$?
  - $n = 10$, $p = 0.5$, $k = 6$
  - $P(X = 6) = \binom{10}{6}(0.5)^6(0.5)^4 = 210 \cdot (0.5)^{10} = 210/1024 \approx 0.205$
- **Worked Example 2:** 20% of items defective. Sample 15 items with replacement. $P(\text{at most 2 defective})$?
  - $X \sim \text{Binomial}(15, 0.2)$
  - $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$
  - $P(X = 0) = \binom{15}{0}(0.2)^0(0.8)^{15} = (0.8)^{15} \approx 0.035$
  - $P(X = 1) = \binom{15}{1}(0.2)^1(0.8)^{14} = 15(0.2)(0.8)^{14} \approx 0.132$
  - $P(X = 2) = \binom{15}{2}(0.2)^2(0.8)^{13} = 105(0.04)(0.8)^{13} \approx 0.231$
  - $P(X \leq 2) \approx 0.398$
- **When to Use Normal Approximation:**
  - Rule of thumb: $np \geq 10$ AND $n(1 - p) \geq 10$
  - Then $X \approx N(np, np(1 - p))$
  - Apply continuity correction: $P(X \leq k) \approx P(Y < k + 0.5)$
- **! NOT Binomial if: sampling without replacement (use Hypergeometric), varying p, dependent trials**
- *Note: Sum of independent Binomials with same p: $Bin(n_1, p) + Bin(n_2, p) = Bin(n_1 + n_2, p)$*

2.3 Poisson Distribution (a.k.a. Poisson(λ), Counting Process)

Terms: λ = rate parameter (avg events per unit time/space), k = number of events observed, Rate = average number of occurrences per unit, Interval = fixed region of time/space, Inter-arrival time = time between consecutive events

- **SYNONYMS:** "arrival process", "counting process", "rare events", "rate λ", "per unit time"
- **Definition:** Count of events in a fixed interval when events occur at constant average rate

– **PMF:** $\boxed{P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots}$

– **Parameters and Moments:**

  * Parameter: $\lambda > 0$ = average rate (events per unit)
  * Mean: $E[X] = \lambda$
  * Variance: $\text{Var}(X) = \lambda$ (mean = variance!)
  * MGF: $M(t) = e^{\lambda(e^t-1)}$

– **Key Properties:**

  * Sum of independent Poissons: $\text{Pois}(\lambda_1) + \text{Pois}(\lambda_2) = \text{Pois}(\lambda_1 + \lambda_2)$
  * Scaling: If rate is λ per hour, rate for t hours is λt
  * Inter-arrival times: $\text{Exp}(\lambda)$ (see Section 3.4)

– **Worked Example 1:** Calls arrive at rate 5 per hour. P(exactly 3 calls in 1 hour)?

  * $X \sim \text{Poisson}(5)$
  * $P(X = 3) = \frac{e^{-5} \cdot 5^3}{3!} = \frac{e^{-5} \cdot 125}{6} \approx 0.140$

– **Worked Example 2:** Same rate. P(at least 2 calls in 30 min)?

  * Rate for 30 min: $\lambda = 5 \times 0.5 = 2.5$
  * $Y \sim \text{Poisson}(2.5)$
  * $P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1)$
  * $P(Y = 0) = e^{-2.5} \approx 0.082$
  * $P(Y = 1) = e^{-2.5}(2.5) \approx 0.205$
  * $P(Y \geq 2) \approx 1 - 0.082 - 0.205 = 0.713$

– **Poisson Approximation to Binomial:**

  * When n large, p small, $\lambda = np$ moderate
  * Rule: $n \geq 20$, $p \leq 0.05$, $np \leq 10$
  * Then $\text{Bin}(n, p) \approx \text{Pois}(np)$

– **Worked Example 3:** 1000 items, each defective with prob 0.002. Approximate P(at least 1 defective).

  * Use Poisson with $\lambda = 1000 \times 0.002 = 2$
  * $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-2} \approx 1 - 0.135 = 0.865$

– *Note: If mean ≠ variance in data, Poisson may not be good fit. Check: $E[X] = Var(X) = \lambda$*

2.4 Geometric Distribution (a.k.a. "First success", Memoryless Discrete)

Terms: p = probability of success on each trial, $q = 1 - p$ = probability of failure, k = count (failures before success OR total trials), Memoryless = past failures don't affect future probability, Tail probability = $P(X \geq k)$

– **SYNONYMS:** "first success", "waiting for success", "trials until success", "how many tries"

– **! TWO CONVENTIONS – know which one is used!**

  * **Our convention:** X = # failures before first success, $k = 0, 1, 2, \ldots$
  * **Alt convention:** Y = # trials until first success, $Y = X + 1$, $k = 1, 2, 3, \ldots$

– **PMF (failures before success):** $\boxed{P(X = k) = p(1-p)^k, \quad k = 0, 1, 2, \ldots}$

– **PMF (trials until success):** $\boxed{P(Y = k) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \ldots}$

– **Parameters (failures convention):**

  * Mean: $E[X] = (1-p)/p = q/p$
  * Variance: $\text{Var}(X) = (1-p)/p^2 = q/p^2$
  * MGF: $M(t) = \frac{p}{1-(1-p)e^t}$ for $t < -\ln(1-p)$

– **Parameters (trials convention):**

  * Mean: $E[Y] = 1/p$
  * Variance: $\text{Var}(Y) = (1-p)/p^2$

– **Memoryless Property (UNIQUE to Geometric among discrete):** $\boxed{P(X > m + n | X > m) = P(X > n)}$

  * "Past failures don't affect future probability"
  * Given you've had m failures, probability of n more is same as starting fresh

– **Useful Tail Probability:** $\boxed{P(X \geq k) = (1-p)^k = q^k}$

– **Worked Example 1:** Roll fair die until first 6. Expected # of rolls?

  * $p = 1/6$ (success = rolling 6)
  * Using trials convention: $E[Y] = 1/p = 6$ rolls

– **Worked Example 2:** Flip biased coin ($P(H) = 0.3$) until first head. P(need exactly 4 flips)?

  * Need 3 tails, then 1 head
  * $P(Y = 4) = (0.7)^3(0.3) = 0.343 \times 0.3 = 0.103$

– **Worked Example 3:** Same as above. Given no head in first 5 flips, expected additional flips?

  * By memoryless property: same as starting fresh!
  * Expected additional = $1/0.3 = 10/3 \approx 3.33$ flips

– *Note: CDF: $P(X \leq k) = 1 - (1-p)^{k+1}$ (failures convention)*

2.5 Negative Binomial

Terms: r = number of successes needed (target), p = probability of success on each trial, k = number of failures before r-th success, Pascal distribution = alternative name

– **SYNONYMS:** "r-th success", "wait for r successes", "Pascal distribution"

– **Definition:** Number of failures before r-th success

– **PMF:** $\boxed{P(X = k) = \binom{k+r-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2}$

  * Interpretation: $\binom{k+r-1}{k}$ = ways to arrange k failures in first $k + r - 1$ trials
  * Last trial must be success (the r-th success)

– **Parameters:**

  * Mean: $E[X] = r(1-p)/p = rq/p$
  * Variance: $\text{Var}(X) = r(1-p)/p^2 = rq/p^2$
  * MGF: $M(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$

– **Relationship to Geometric:**

  * Geometric = Negative Binomial with $r = 1$
  * Sum of r i.i.d. Geometric(p) = Negative Binomial(r, p)

– **Worked Example:** Flip coin ($p = 0.4$). P(exactly 5 failures before 3rd success)?

  * $X \sim \text{NegBin}(3, 0.4)$, find $P(X = 5)$
  * $P(X = 5) = \binom{5+3-1}{5}(0.4)^3(0.6)^5 = \binom{7}{5}(0.064)(0.07776)$

* $= 21 \times 0.064 \times 0.07776 \approx 0.104$
- **Alternative (trials) convention:**
  * $Y =$ trials until r-th success, $Y = X + r$
  * Mean: $E[Y] = r/p$
- *Note: Useful for quality control: "How many items until r defectives?"*

2.6 Hypergeometric Distribution

Terms: N = population size (total items), K = number of "success" items in population, n = sample size (items drawn), k = successes in sample, Without replacement = once drawn, item cannot be drawn again

- **SYNONYMS:** "without replacement", "finite population sampling", "lottery"
- **Definition:** Number of successes when sampling n items without replacement from population of N items containing K successes
- **PMF:** $\boxed{P(X = k) = \dfrac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}}$
  * $\binom{K}{k}$ = ways to choose k successes from K available
  * $\binom{N-K}{n-k}$ = ways to choose $n-k$ failures from $N-K$ available
  * $\binom{N}{n}$ = total ways to choose n from N
- **Parameters:**
  * N = population size
  * K = number of successes in population
  * n = sample size
  * k = observed successes in sample (what we count)
- **Valid Range:** $\max(0, n-(N-K)) \leq k \leq \min(n, K)$
- **Moments:**
  * Mean: $E[X] = n \cdot \frac{K}{N}$
  * Variance: $\text{Var}(X) = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$
  * Note: Variance has finite population correction factor $\frac{N-n}{N-1}$
- **Worked Example 1:** Deck of 52 cards. Draw 5 cards. $P(\text{exactly 2 aces})$?
  * $N = 52$, $K = 4$ (aces), $n = 5$, $k = 2$
  * $P(X = 2) = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} = \frac{6 \times 17296}{2598960} \approx 0.040$
- **Worked Example 2:** Box: 10 red, 15 blue balls. Draw 6 without replacement. $P(\text{at least 4 red})$?
  * $N = 25$, $K = 10$, $n = 6$
  * $P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6)$
  * $P(X = 4) = \frac{\binom{10}{4}\binom{15}{2}}{\binom{25}{6}} = \frac{210 \times 105}{177100} \approx 0.124$
  * Continue for $k = 5, 6$ and sum
- **Hypergeometric vs Binomial:**
  * Hypergeometric: without replacement, trials dependent
  * Binomial: with replacement (or large population), trials independent
  * Approximation: If $n/N < 0.05$, Hypergeometric $\approx$ Binomial$(n, K/N)$
- **! Use Hypergeometric when: finite population, no replacement, population size matters**
- *Note: Hypergeometric mean = Binomial mean (same formula), but variance is smaller due to finite population correction*

3. CONTINUOUS RANDOM VARIABLES

3.1 PDF and CDF

Terms: PDF = Probability Density Function $f(x)$, CDF = Cumulative Distribution Function $F(x) = P(X \leq x)$, $E[X]$ = expected value, $\text{Var}(X)$ = variance, LOTUS = Law of the Unconscious Statistician, Quantile x_p = value where $F(x_p) = p$

- **PDF (Probability Density Function):**
  * $f(x) \geq 0$ for all x
  * $\int_{-\infty}^{\infty} f(x)dx = 1$ (normalization)
  * NOT a probability! $f(x)$ can be > 1
  * $f(x)dx \approx P(x < X < x + dx)$ (infinitesimal probability)
- **CDF (Cumulative Distribution Function):**

$$\boxed{F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt}$$

  * Properties: $0 \leq F(x) \leq 1$, non-decreasing, continuous
  * $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$
- **Key Relationships:**
  * PDF from CDF: $f(x) = F'(x)$ (derivative)
  * CDF from PDF: $F(x) = \int_{-\infty}^{x} f(t)dt$
- **Probability Calculations:**
  * $\boxed{P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)}$
  * $P(X > a) = 1 - F(a)$
  * $P(X < a) = F(a)$ (same as $P(X \leq a)$ for continuous!)
  * **! $P(X = a) = 0$ for any specific value a!**
- **Expectation and Variance:**
  * $\boxed{E[X] = \int_{-\infty}^{\infty} x f(x)dx}$
  * $\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx}$ (LOTUS)
  * $\boxed{\text{Var}(X) = E[X^2] - (E[X])^2}$ (computational formula - USE THIS!)
  * $\boxed{\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx}$ (definition)
- **Worked Example:** $f(x) = 2x$ for $0 \leq x \leq 1$, else 0.
  * Verify: $\int_0^1 2x\,dx = [x^2]_0^1 = 1$ ✓
  * CDF: $F(x) = \int_0^x 2t\,dt = x^2$ for $0 \leq x \leq 1$
  * $P(0.5 < X < 0.8) = F(0.8) - F(0.5) = 0.64 - 0.25 = 0.39$
  * $E[X] = \int_0^1 x \cdot 2x\,dx = \int_0^1 2x^2\,dx = [2x^3/3]_0^1 = 2/3$
  * $E[X^2] = \int_0^1 x^2 \cdot 2x\,dx = \int_0^1 2x^3\,dx = [x^4/2]_0^1 = 1/2$

* $\text{Var}(X) = 1/2 - (2/3)^2 = 1/2 - 4/9 = 1/18$

- **Finding Constant c Procedure:**
  1. Given $f(x) = c \cdot g(x)$ on some region
  2. Set $\int_{\text{region}} c \cdot g(x)\,dx = 1$
  3. Solve for c
- *Note: Quantile: x_p where $F(x_p) = p$. Median: $x_{0.5}$, Quartiles: $x_{0.25}$, $x_{0.75}$*

3.2 Uniform Distribution

Terms: a = lower bound, b = upper bound, $U(a,b)$ = Uniform on interval $[a,b]$, Standard Uniform = $U(0,1)$, Maximum entropy = most "spread out" distribution given constraints

- **SYNONYMS:** "equally likely", "random point", "uniform random"
- **Definition:** All values in $[a,b]$ equally likely
- **Notation:** $X \sim \text{Uniform}(a,b)$ or $X \sim U(a,b)$
- **PDF:** $\boxed{f(x) = \frac{1}{b-a}, \quad a \le x \le b}$
  * $f(x) = 0$ for $x < a$ or $x > b$
- **CDF:** $\boxed{F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}}$
- **Parameters:**
  * Mean: $E[X] = \frac{a+b}{2}$ (midpoint)
  * Variance: $\text{Var}(X) = \frac{(b-a)^2}{12}$
  * MGF: $M(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$
- **Key Property:** $\boxed{P(c < X < d) = \frac{d-c}{b-a}}$
  * Probability is proportional to length of interval!
- **Standard Uniform** $U(0,1)$**:**
  * $f(x) = 1$ for $0 \le x \le 1$
  * $F(x) = x$ for $0 \le x \le 1$
  * $E[X] = 0.5$, $\text{Var}(X) = 1/12$
  * Used in simulation: $F^{-1}(U)$ generates random variable with CDF F
- **Worked Example 1:** Bus arrives uniformly between 8:00-8:30. $P(\text{wait} < 10 \text{ min})$ if you arrive at 8:15?
  * If bus arrives before 8:15, you missed it (assume you want wait < 10 min)
  * Actually: Bus arrival time $\sim U(0, 30)$ min after 8:00
  * You arrive at 15 min. You wait < 10 min if bus arrives in (15, 25)
  * $P(15 < X < 25) = (25-15)/30 = 10/30 = 1/3$
- **Worked Example 2:** $X \sim U(2,8)$. Find $P(X > 5)$ and $E[X^2]$.
  * $P(X > 5) = (8-5)/(8-2) = 3/6 = 1/2$
  * $E[X] = (2+8)/2 = 5$
  * $\text{Var}(X) = (8-2)^2/12 = 36/12 = 3$
  * $E[X^2] = \text{Var}(X) + (E[X])^2 = 3 + 25 = 28$
- **Linear Transformation:**
  * If $U \sim U(0,1)$, then $X = a + (b-a)U \sim U(a,b)$
  * Conversely: $U = (X-a)/(b-a) \sim U(0,1)$
- *Note: Uniform is the "maximum entropy" distribution on a bounded interval*

3.3 Normal Distribution (a.k.a. Gaussian, N(μ,σ^2))

* **High Priority!** **Terms:** μ = mean (center), σ^2 = variance (spread), σ = standard deviation, Z = standard normal $N(0,1)$, $\Phi(z)$ = CDF of standard normal, $\phi(z)$ = PDF of standard normal, Standardization = convert to $Z = (X-\mu)/\sigma$

- **SYNONYMS: Gaussian = Normal =** N(μ,σ^2) = "bell curve"
- **! "Gaussian" means Normal! Don't be confused by terminology!**
- **PDF:** $\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty}$
- **Notation:** $X \sim N(\mu, \sigma^2)$
  * μ = mean (center of distribution)
  * σ^2 = variance (spread), σ = standard deviation
  * **! $N(\mu,\sigma^2)$ uses VARIANCE, not std dev!**
- **Standard Normal** $Z \sim N(0,1)$**:**
  * PDF: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
  * CDF: $\Phi(z) = P(Z \le z)$ (use table)
  * Symmetry: $\Phi(-z) = 1 - \Phi(z)$
- **Standardization (CRITICAL SKILL):** $\boxed{Z = \frac{X-\mu}{\sigma} \sim N(0,1)}$
  * Convert any normal to standard normal
  * De-standardize: $X = \mu + \sigma Z$
- **Step-by-Step: Finding** $P(X < a)$**:**
  1. Standardize: $z = \frac{a-\mu}{\sigma}$
  2. Look up $\Phi(z)$ in table
  3. That's your answer!
- **Step-by-Step: Finding** $P(a < X < b)$**:**
  1. Standardize both: $z_a = \frac{a-\mu}{\sigma}$, $z_b = \frac{b-\mu}{\sigma}$
  2. $P(a < X < b) = \Phi(z_b) - \Phi(z_a)$
- **Worked Example 1:** $X \sim N(100, 225)$. Find $P(X < 115)$.
  * Note: $\sigma^2 = 225$, so $\sigma = 15$
  * $z = (115-100)/15 = 1$
  * $P(X < 115) = \Phi(1) = 0.8413$
- **Worked Example 2:** $X \sim N(50, 16)$. Find $P(45 < X < 55)$.
  * $\sigma = 4$
  * $z_1 = (45-50)/4 = -1.25$, $z_2 = (55-50)/4 = 1.25$
  * $P(45 < X < 55) = \Phi(1.25) - \Phi(-1.25) = 0.8944 - 0.1056 = 0.7888$
- **Worked Example 3:** $X \sim N(70, 100)$. Find c such that $P(X > c) = 0.05$.
  * Need $P(X \le c) = 0.95$
  * From table: $z = 1.645$ gives $\Phi(z) = 0.95$
  * $c = \mu + z\sigma = 70 + 1.645(10) = 86.45$
- **Linear Transformation Property:** $\boxed{\text{If } X \sim N(\mu,\sigma^2), \text{ then } aX + b \sim N(a\mu + b, a^2\sigma^2)}$
- **Sum of Independent Normals:** $\boxed{X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)}$

- More generally: $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$
- **! Requires independence! For dependent, see Sec 4.5**

- **68-95-99.7 Rule:**
  - $P(\mu - \sigma < X < \mu + \sigma) \approx 0.68$
  - $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$
  - $P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997$
- **MGF:** $M(t) = e^{\mu t + \sigma^2 t^2/2}$
- *Note: If $X \sim N(\mu, \sigma^2)$, then $E[X] = \mu$, $E[X^2] = \sigma^2 + \mu^2$*

3.4 Exponential Distribution (a.k.a. Exp(λ), Memoryless)

Terms: λ = rate parameter (events per unit time), $1/\lambda$ = mean (avg waiting time), s, t = time values in memoryless property, Memoryless = $P(X > s + t|X > s) = P(X > t)$, Survival function = $P(X > x) = e^{-\lambda x}$, Inter-arrival time = time between consecutive Poisson events

- **SYNONYMS:** Exp(λ), "waiting time", "memoryless", "inter-arrival time", "lifetime"
- **! PARAMETER TRAP – Know how to convert!**
  - **"Rate $\lambda = 3$"** means $\lambda = 3$, Mean $= 1/3$
  - **"Mean $\theta = 3$"** means Mean $= 3$, so $\lambda = 1/3$
  - Formula: $\lambda = 1/\text{mean}$, Mean $= 1/\lambda$
- **PDF:** $\boxed{f(x) = \lambda e^{-\lambda x}, \quad x > 0}$
- **CDF:** $\boxed{F(x) = 1 - e^{-\lambda x}, \quad x \geq 0}$
- **Survival Function:** $P(X > x) = e^{-\lambda x}$
- **Parameters:**
  - Mean: $E[X] = 1/\lambda$
  - Variance: $\text{Var}(X) = 1/\lambda^2$
  - Median: $\ln(2)/\lambda \approx 0.693/\lambda$
  - MGF: $M(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$
- **Memoryless Property (UNIQUE to Exponential among continuous):** $\boxed{P(X > s + t\,|\,X > s) = P(X > t)}$
  - "Given survived s, additional survival is same as starting fresh"
  - Proof: $\frac{P(X>s+t)}{P(X>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$
- **Relationship to Poisson:**
  - If events arrive as Poisson(λ), time between events is Exp(λ)
  - If $X \sim$ Exp(λ) = time until event, then # events in time t is Poisson(λt)
- **Worked Example 1:** Service time Exp with mean 5 min. $P(\text{service} > 8 \text{ min})$?
  - Mean = 5, so $\lambda = 1/5 = 0.2$
  - $P(X > 8) = e^{-0.2\times 8} = e^{-1.6} \approx 0.202$
- **Worked Example 2:** Same service. Given waited 3 min, $P(\text{wait at least 5 more min})$?
  - By memoryless property: $P(X > 8|X > 3) = P(X > 5)$
  - $P(X > 5) = e^{-0.2\times 5} = e^{-1} \approx 0.368$
- **Worked Example 3:** Component lifetime Exp($\lambda = 0.01$). Expected lifetime? $P(\text{survives} > 200 \text{ hours})$?
  - $E[X] = 1/0.01 = 100$ hours
  - $P(X > 200) = e^{-0.01\times 200} = e^{-2} \approx 0.135$
- **Minimum of Independent Exponentials:** $\boxed{\min(X_1, ..., X_n) \sim \text{Exp}(\lambda_1 + \lambda_2 + ... + \lambda_n)}$
  - Proof: $P(\min > t) = P(\text{all} > t) = \prod e^{-\lambda_i t} = e^{-(\sum \lambda_i)t}$
- **Sum of Exponentials:**
  - $\sum_{i=1}^{n}$ Exp(λ) (same rate) $\sim$ Gamma(n, λ)
- *Note: Hazard rate is constant: $h(x) = f(x)/(1 - F(x)) = \lambda$. This defines memorylessness!*

3.5 Gamma Distribution (a.k.a. Gamma(r, λ), Erlang)

Terms: r = shape parameter (# of events to wait for), λ = rate parameter, $\Gamma(r)$ = Gamma function (generalizes factorial), Erlang = Gamma with integer r, Chi-square = special Gamma with $r = n/2$, $\lambda = 1/2$

- **SYNONYMS:** "sum of exponentials", "time until r-th event", Erlang (integer r)
- **Definition:** Generalization of Exponential; time until r-th Poisson event
- **PDF:** $\boxed{f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0}$
- **Parameters:**
  - $r > 0$ = shape parameter
  - $\lambda > 0$ = rate parameter
  - Mean: $E[X] = r/\lambda$
  - Variance: $\text{Var}(X) = r/\lambda^2$
  - MGF: $M(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$ for $t < \lambda$
- **Gamma Function:**
  - $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$
  - $\Gamma(n) = (n - 1)!$ for positive integer n
  - $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$
  - Recursion: $\Gamma(r + 1) = r \cdot \Gamma(r)$
- **Special Cases:**
  - $r = 1$: Gamma($1, \lambda$) = Exponential(λ)
  - $r = n/2$, $\lambda = 1/2$: Chi-square with n degrees of freedom
  - Integer r: Called Erlang distribution
- **Key Property:** $\boxed{\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda) \text{ if } X_i \overset{iid}{\sim} \text{Exp}(\lambda}$
- **Worked Example:** Time for 5 customers to arrive if arrivals are Poisson with rate 2/hour.
  - Time until 5th arrival $\sim$ Gamma(5, 2)
  - $E[T] = 5/2 = 2.5$ hours
  - $\text{Var}(T) = 5/4 = 1.25$ hours^2
- **Sum of Gammas (same λ):** $\boxed{\text{Gamma}(r_1, \lambda) + \text{Gamma}(r_2, \lambda) = \text{Gamma}(r_1 + r_2, \lambda)}$
- *Note: Gamma is conjugate prior for Poisson rate parameter (Bayesian)*

3.6 Beta Distribution (a.k.a. Beta(α, β), Conjugate Prior)

Terms: α = shape parameter ("prior successes" + 1), β = shape parameter ("prior failures" + 1), $B(\alpha,\beta)$ = Beta function (normalizing constant), Conjugate prior = posterior is same distribution family as prior, Jeffreys prior = "uninformative" prior

- **SYNONYMS:** "prior for probability", "proportion model", "conjugate to Binomial"
- **Definition:** Models probabilities/proportions on $(0, 1)$
- **PDF:** $\boxed{f(x) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1}$
- **Alternative notation:** $f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
  - where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$
- **Parameters:**
  - $\alpha > 0$, $\beta > 0$ = shape parameters
  - Mean: $E[X] = \frac{\alpha}{\alpha+\beta}$
  - Mode: $\frac{\alpha-1}{\alpha+\beta-2}$ (for $\alpha, \beta > 1$)
  - Variance: $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- **Special Cases:**
  - $\alpha = \beta = 1$: Uniform$(0, 1)$
  - $\alpha = \beta$: Symmetric around 0.5
  - $\alpha > \beta$: Skewed right (mode > 0.5)
  - $\alpha < \beta$: Skewed left (mode < 0.5)
  - $\alpha = \beta = 0.5$: Arcsine distribution
- **Shape Interpretation:**
  - $\alpha - 1$ = "prior successes"
  - $\beta - 1$ = "prior failures"
  - Large $\alpha + \beta$ = more concentrated around mean
- **Conjugate Prior for Binomial (CRITICAL for Bayesian):**
  - Prior: $p \sim \text{Beta}(\alpha, \beta)$
  - Data: k successes in n trials (Binomial)
  - Posterior: $p|\text{data} \sim \text{Beta}(\alpha + k, \beta + n - k)$
  - "Add successes to α, failures to β"
- **Worked Example:** Prior belief: $p \sim \text{Beta}(2, 2)$. Observe 3 successes in 5 trials.
  - Posterior: $p|\text{data} \sim \text{Beta}(2 + 3, 2 + 2) = \text{Beta}(5, 4)$
  - Prior mean: $2/(2 + 2) = 0.5$
  - Posterior mean: $5/(5 + 4) = 5/9 \approx 0.556$
- **Jeffreys Prior:** $\text{Beta}(1/2, 1/2)$ = "uninformative" prior for probability
- *Note: Beta function: $B(\alpha,\beta) = B(\beta,\alpha)$ (symmetric)*

4. MULTIVARIATE DISTRIBUTIONS

*** Post-M2**

4.1 Joint Distributions

Terms: Joint PMF/PDF = $f(x, y)$ describes both variables together, Joint CDF = $F(x, y) = P(X \le x, Y \le y)$, Support = region where $f(x, y) > 0$, Double integral = integrate over 2D region

- **Joint PMF (Discrete):**
  - $p(x, y) = P(X = x, Y = y)$
  - Requirements: $p(x, y) \ge 0$, $\sum_x \sum_y p(x, y) = 1$
- **Joint PDF (Continuous):**
  - $f(x, y) \ge 0$ and $\int\int f(x, y)dxdy = 1$
  - $P((X, Y) \in A) = \iint_A f(x, y)dxdy$
- **Joint CDF:** $\boxed{F(x, y) = P(X \le x, Y \le y)}$
  - Continuous: $F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s, t)\,dt\,ds$
  - PDF from CDF: $f(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$
- **Procedure: Finding Constant c:**
  1. Set up double integral over support region
  2. Integrate: $\int\int c \cdot g(x, y)\,dx\,dy = 1$
  3. Solve for c
- **Worked Example 1:** $f(x, y) = c(x^2 + xy)$ on $[0, 1] \times [0, 1]$. Find c.
  - $\int_0^1 \int_0^1 c(x^2 + xy)\,dx\,dy = 1$
  - Inner integral: $\int_0^1 (x^2 + xy)\,dx = [\frac{x^3}{3} + \frac{x^2 y}{2}]_0^1 = \frac{1}{3} + \frac{y}{2}$
  - Outer integral: $\int_0^1 (\frac{1}{3} + \frac{y}{2})\,dy = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$
  - So $c \cdot \frac{7}{12} = 1 \Rightarrow c = \frac{12}{7}$
- **Worked Example 2:** $f(x, y) = c$ on triangle $0 < x < y < 1$. Find c.
  - Region: $0 < x < 1$, $x < y < 1$
  - $\int_0^1 \int_x^1 c\,dy\,dx = c\int_0^1 (1 - x)\,dx = c \cdot \frac{1}{2} = 1$
  - $c = 2$
- **Probability Over Region:**
  1. Identify region A carefully (draw it!)
  2. Set up double integral with correct limits
  3. Integrate: $P((X, Y) \in A) = \iint_A f(x, y)\,dx\,dy$
- **Common Integration Regions:**
  - Rectangle: $\int_a^b \int_c^d f(x, y)\,dy\,dx$
  - Triangle ($y > x$): $\int_0^1 \int_x^1 f(x, y)\,dy\,dx$
  - Circle: Convert to polar coordinates
- **! Always draw the region! Check integration limits match the support!**
- *Note: Order of integration can be switched (Fubini's theorem) if integral converges*

4.2 Marginal and Conditional Distributions

Terms: Marginal = distribution of one variable alone (integrate/sum out the other), $f_X(x)$ = marginal PDF of X, Conditional = $f_{Y|X}(y|x)$ = PDF of Y given $X = x$, "Integrate out" = integrate over all values of unwanted variable

- **Marginal Distributions – "Integrate out" the other variable:**
  - $\boxed{f_X(x) = \displaystyle\int_{-\infty}^{\infty} f(x, y)\,dy}$ (continuous)
  - $\boxed{p_X(x) = \displaystyle\sum_y p(x, y)}$ (discrete)
  - Similarly: $f_Y(y) = \int f(x, y)\,dx$

- **Conditional PDF:** $\boxed{f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad \text{provided } f_X(x) > 0}$
  - This is a valid PDF in y (integrates to 1 in y)
  - Read as: "density of Y given $X = x$"
- **Step-by-Step: Finding Marginal $f_X(x)$:**
  1. Look at joint PDF support region
  2. For fixed x, determine what values of y are allowed
  3. Integrate joint over those y values: $f_X(x) = \int_{y_{\min}(x)}^{y_{\max}(x)} f(x,y)\,dy$
  4. State the support of $f_X(x)$
- **Worked Example:** Joint: $f(x,y) = 2$ on $0 < x < y < 1$.
  - **Marginal of X:** For fixed $x \in (0,1)$, y ranges from x to 1
    - $f_X(x) = \int_x^1 2\,dy = 2(1-x)$ for $0 < x < 1$
  - **Marginal of Y:** For fixed $y \in (0,1)$, x ranges from 0 to y
    - $f_Y(y) = \int_0^y 2\,dx = 2y$ for $0 < y < 1$
  - **Conditional $Y|X = x$:**
    - $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x}$ for $x < y < 1$
    - This is Uniform$(x, 1)$!
  - **Conditional $X|Y = y$:**
    - $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{2}{2y} = \frac{1}{y}$ for $0 < x < y$
    - This is Uniform$(0, y)$!
- **Worked Example 2:** $f(x,y) = \frac{12}{7}(x^2 + xy)$ on $[0,1]^2$. Find $f_X(x)$.
  - $f_X(x) = \int_0^1 \frac{12}{7}(x^2+xy)\,dy = \frac{12}{7}[x^2 y + \frac{xy^2}{2}]_0^1 = \frac{12}{7}(x^2 + \frac{x}{2})$
  - $f_X(x) = \frac{12}{7}x(x + \frac{1}{2}) = \frac{6x(2x+1)}{7}$ for $0 < x < 1$
- **Key Properties:**
  - $\int f_{Y|X}(y|x)\,dy = 1$ (valid density)
  - $f(x,y) = f_{Y|X}(y|x) \cdot f_X(x) = f_{X|Y}(x|y) \cdot f_Y(y)$
- **! Bounds in marginal integral depend on the support region – draw it!**
- *Note: Conditional expectation:* $E[Y|X = x] = \int y \cdot f_{Y|X}(y|x)\,dy$

4.3 Independence of Random Variables

Terms: Independent $=$ $f(x,y) = f_X(x) \cdot f_Y(y)$ for all x, y, i.i.d. $=$ independent and identically distributed, Uncorrelated $=$ $\text{Cov}(X,Y) = 0$ (weaker than independence), Dependent $=$ not independent

- **Definition:** X, Y independent iff $\boxed{f(x,y) = f_X(x) \cdot f_Y(y) \text{ for ALL } x, y}$
- **Equivalent Conditions:**
  - Joint $=$ product of marginals
  - $f_{Y|X}(y|x) = f_Y(y)$ (conditioning doesn't change distribution)
  - $F(x,y) = F_X(x) \cdot F_Y(y)$
- **Step-by-Step: Testing Independence:**
  1. Find joint PDF/PMF $f(x,y)$
  2. Find marginal $f_X(x)$ by integrating out y
  3. Find marginal $f_Y(y)$ by integrating out x
  4. Check: Does $f(x,y) = f_X(x) \cdot f_Y(y)$ for ALL (x,y) in support?
  5. If YES at all points: Independent. If NO anywhere: Dependent.
- **Quick Check for Independence:**
  - Can $f(x,y)$ be written as $g(x) \cdot h(y)$?
  - AND support must be a rectangle (product of intervals)?
  - If both YES $\Rightarrow$ likely independent (verify by checking)
- **Worked Example 1:** $f(x,y) = e^{-x-y}$ for $x > 0$, $y > 0$.
  - Can factor: $f(x,y) = e^{-x} \cdot e^{-y}$
  - Support: $(0,\infty) \times (0,\infty) =$ rectangle ✓
  - Marginals: $f_X(x) = e^{-x}$, $f_Y(y) = e^{-y}$ (both Exp(1))
  - Check: $f_X(x) \cdot f_Y(y) = e^{-x} \cdot e^{-y} = e^{-x-y} = f(x,y)$ ✓
  - **Independent!**
- **Worked Example 2:** $f(x,y) = 2$ for $0 < x < y < 1$.
  - Support: Triangle (NOT a rectangle!)
  - Therefore: **NOT independent!** (no need to compute marginals)
  - Intuition: Knowing X constrains possible values of Y
- **Worked Example 3:** $f(x,y) = \frac{12}{7}(x^2 + xy)$ on $[0,1]^2$.
  - Support is rectangle, but can we factor?
  - $x^2 + xy = x(x+y)$ – cannot write as $g(x)h(y)$
  - **NOT independent!**
- **Consequences of Independence:**
  - $E[XY] = E[X] \cdot E[Y]$
  - $E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$
  - $\text{Cov}(X,Y) = 0$
  - $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
  - $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$
- **! $\text{Cov}(X,Y) = 0$ does NOT imply independence!**
  - Counterexample: $X \sim N(0,1)$, $Y = X^2$. Then $\text{Cov}(X,Y) = E[X^3] = 0$, but clearly dependent!
  - **EXCEPTION:** For jointly normal (bivariate normal), $\rho = 0 \Leftrightarrow$ independent!
- *Note: i.i.d. = independent AND identically distributed*

4.4 Covariance and Correlation

Terms: $\text{Cov}(X,Y)$ = covariance (measures linear relationship), ρ_{XY} = correlation (unitless, $-1 \le \rho \le 1$), σ_X = standard deviation of X, $E[XY]$ = expected value of product, Uncorrelated = $\rho = 0$

- **Covariance Definition:** $\boxed{\text{Cov}(X,Y) = E[(X - \mu_X)(Y}$
- **Computational Formula (USE THIS!):** $\boxed{\text{Cov}(X,Y) = E[XY] - E[X]E[Y]}$
- **Correlation:** $\boxed{\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho \le 1}$
  - $\rho = 1$: Perfect positive linear relationship
  - $\rho = -1$: Perfect negative linear relationship
  - $\rho = 0$: No linear relationship (uncorrelated)
- **Step-by-Step: Computing Covariance:**
  1. Find $E[X] = \int\int x \cdot f(x,y)\,dx\,dy$

2. Find $E[Y] = \int\int y \cdot f(x,y)\,dx\,dy$
3. Find $E[XY] = \int\int xy \cdot f(x,y)\,dx\,dy$
4. Compute: $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$

– **Step-by-Step: Computing Correlation:**

1. Find $\text{Cov}(X,Y)$ (steps above)
2. Find $E[X^2]$, compute $\text{Var}(X) = E[X^2] - (E[X])^2$, then $\sigma_X = \sqrt{\text{Var}(X)}$
3. Similarly find σ_Y
4. Compute: $\rho = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$

– **Key Properties of Covariance:**

* $\text{Cov}(X,X) = \text{Var}(X)$
* $\text{Cov}(X,Y) = \text{Cov}(Y,X)$ (symmetric)
* $\text{Cov}(aX+b, cY+d) = ac \cdot \text{Cov}(X,Y)$
* $\text{Cov}(X+Y,Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$
* If independent: $\text{Cov}(X,Y) = 0$

– **Variance of Sum:** $\boxed{\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)}$

$\boxed{\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y)}$

$\boxed{\text{Var}(aX+bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)}$

– **Worked Example:** $f(x,y) = \frac{12}{7}(x^2+xy)$ on $[0,1]^2$. Find ρ.

* $E[X] = \frac{12}{7}\int_0^1\int_0^1 x(x^2+xy)\,dy\,dx = \frac{12}{7}\cdot\frac{5}{12} = \frac{5}{7}$
* $E[Y] = \frac{12}{7}\int_0^1\int_0^1 y(x^2+xy)\,dx\,dy = \frac{12}{7}\cdot\frac{11}{24} = \frac{11}{14}$
* $E[XY] = \frac{12}{7}\int_0^1\int_0^1 xy(x^2+xy)\,dx\,dy = \frac{12}{7}\cdot\frac{5}{12} = \frac{5}{7}$
* $\text{Cov}(X,Y) = \frac{5}{7} - \frac{5}{7}\cdot\frac{11}{14} = \frac{5}{7}(1-\frac{11}{14}) = \frac{5}{7}\cdot\frac{3}{14} = \frac{15}{98}$
* Then find variances and compute ρ

– **[$] Portfolio: Var**$(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y$

– *Note: Correlation is unitless and scale-invariant:* $\rho_{aX+b,cY+d} = \text{sign}(ac)\cdot\rho_{XY}$

4.5 Bivariate Normal (a.k.a. Gaussian Vector, MVN, Jointly Normal)

* **Critical! Terms:** $\boldsymbol{\mu}$ = mean vector, $\boldsymbol{\Sigma}$ = covariance matrix, ρ = correlation between X and Y, MVN = Multivariate Normal, Gaussian = Normal, Jointly Normal = both variables follow bivariate normal together, Independent components = $\rho = 0$ (diagonal $\boldsymbol{\Sigma}$)

– **SYNONYMS: Gaussian vector = MVN = Multivariate Normal** = "Jointly Normal"

– **! KEY TERMINOLOGY:**

* **"Gaussian"** = Normal!
* **"Gaussian vector"** = Multivariate Normal!
* **"Independent components"** = $\rho = 0$ = **INDEPENDENT** (for MVN only!)

– **5 Parameters:** $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$

MEAN VECTOR AND COVARIANCE MATRIX:

– **Mean Vector:** $\boxed{\boldsymbol{\mu} = \begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix} = \begin{pmatrix}E[X]\\ E[Y]\end{pmatrix}}$

Covariance Matrix: $\boxed{\boldsymbol{\Sigma} = \begin{pmatrix}\sigma_X^2 & \text{Cov}(X,Y)\\ \text{Cov}(X,Y) & \sigma_Y^2\end{pmatrix} = \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}}$

– **Procedure: Construct Covariance Matrix from Parameters:**

1. Given: σ_X^2, σ_Y^2, ρ
2. Compute off-diagonal: $\text{Cov}(X,Y) = \rho\cdot\sigma_X\cdot\sigma_Y$
3. Fill matrix: diagonal = variances, off-diagonal = covariance

– **Worked Example:** $X \sim N(1,4)$, $Y \sim N(-2,9)$, $\rho = -0.5$

* $\sigma_X = 2$, $\sigma_Y = 3$
* $\text{Cov}(X,Y) = (-0.5)(2)(3) = -3$
* $\boldsymbol{\mu} = \begin{pmatrix}1\\ -2\end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{pmatrix}4 & -3\\ -3 & 9\end{pmatrix}$

INDEPENDENT COMPONENTS (Critical!):

– **For MVN ONLY:** $\rho = 0 \Leftrightarrow X, Y$ independent

– **Covariance Matrix for Independent:** $\boxed{\boldsymbol{\Sigma} = \begin{pmatrix}\sigma_X^2 & 0\\ 0 & \sigma_Y^2\end{pmatrix}}$ (diagonal!)

– **Procedure: Create Gaussian Vector with Independent Components:**

1. Start with $X_1, X_2 \overset{iid}{\sim} N(0,1)$
2. Define $Y_1 = aX_1 + bX_2$, $Y_2 = cX_1 + dX_2$
3. For independence: Set $\text{Cov}(Y_1,Y_2) = 0$
4. $\text{Cov}(Y_1,Y_2) = ac\cdot\text{Var}(X_1) + bd\cdot\text{Var}(X_2) = ac + bd$
5. Solve $ac + bd = 0$ for desired relationship
6. Result: (Y_1, Y_2) is Gaussian vector with independent components

– **Worked Example:** $Y_1 = aX_1 + X_2$, $Y_2 = X_1 + bX_2$ where $X_1, X_2 \overset{iid}{\sim} N(0,1)$.

* Find b such that Y_1, Y_2 are independent.
* $\text{Cov}(Y_1,Y_2) = a\cdot 1\cdot 1 + 1\cdot b\cdot 1 = a + b$
* For independence: $a + b = 0 \Rightarrow b = -a$
* Marginals: $Y_1 \sim N(0, a^2+1)$, $Y_2 \sim N(0, 1+a^2)$
* Joint density: $f(y_1,y_2) = f_{Y_1}(y_1)\cdot f_{Y_2}(y_2)$ (product of marginals)

LINEAR COMBINATIONS:

– **Key Property:** Any linear combination of jointly normal RVs is normal!

– **For** $Z = aX + bY$**:**

* $\boxed{E[Z] = a\mu_X + b\mu_Y}$
* $\boxed{\text{Var}(Z) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y}$
* $\boxed{Z \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)}$

– **Step-by-Step: Find** $P(aX + bY > c)$**:**

1. Let $Z = aX + bY$
2. Compute $\mu_Z = a\mu_X + b\mu_Y$
3. Compute $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y$
4. $Z \sim N(\mu_Z, \sigma_Z^2)$
5. Standardize: $P(Z > c) = P\left(\frac{Z-\mu_Z}{\sigma_Z} > \frac{c-\mu_Z}{\sigma_Z}\right) = 1 - \Phi\left(\frac{c-\mu_Z}{\sigma_Z}\right)$

– **Worked Example:** $X \sim N(1,2)$, $Y \sim N(-2,3)$, $\rho = -2/3$. Find $P(X+Y>0)$.

* $Z = X + Y$: $a = b = 1$
* $\mu_Z = 1 + (-2) = -1$
* $\sigma_X = \sqrt{2}$, $\sigma_Y = \sqrt{3}$
* $\sigma_Z^2 = 2 + 3 + 2(1)(1)(-2/3)(\sqrt{2})(\sqrt{3}) = 5 - \frac{4\sqrt{6}}{3} \approx 1.73$

* $P(Z > 0) = 1 - \Phi\left(\frac{0-(-1)}{\sqrt{1.73}}\right) = 1 - \Phi(0.76) \approx 0.224$

CONDITIONAL DISTRIBUTIONS:

- **Conditional Distribution $Y|X = x$:** $\boxed{Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1-\rho^2)\right)}$
- **Conditional Mean (regression line):** $\boxed{E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)}$
- **Conditional Variance (constant!):** $\boxed{\text{Var}(Y|X) = \sigma_Y^2(1-\rho^2)}$
  * Does NOT depend on x! Always the same.
  * When $|\rho| = 1$: Variance $= 0$ (perfect prediction)
  * When $\rho = 0$: Variance $= \sigma_Y^2$ (no information from X)
- **Worked Example:** $X \sim N(0,1)$, $Y \sim N(0,4)$, $\rho = 0.8$. Find $E[Y|X = 2]$ and $\text{Var}(Y|X)$.
  * $E[Y|X = 2] = 0 + 0.8 \cdot \frac{2}{1} \cdot (2 - 0) = 3.2$
  * $\text{Var}(Y|X) = 4(1 - 0.64) = 4(0.36) = 1.44$
  * $Y|X = 2 \sim N(3.2, 1.44)$
- **Special Case:** $\rho = 0$:
  * $E[Y|X = x] = \mu_Y$ (doesn't depend on x!)
  * $\text{Var}(Y|X) = \sigma_Y^2$
  * Conditional = Marginal (by independence)

Note: For BVN: uncorrelated $\Leftrightarrow$ independent. This is UNIQUE to normal distributions!

4.6 Transformations (a.k.a. Jacobian Method, CDF Method)

* Complex! **Terms:** Transformation $= Y = g(X)$, Jacobian $J =$ determinant of matrix of partial derivatives, CDF method = find CDF then differentiate, Monotonic = strictly increasing or decreasing, g^{-1} = inverse function

- **SYNONYMS:** "change of variables", "find distribution of $Y = g(X)$", "Jacobian"

SINGLE VARIABLE TRANSFORMATIONS:

- **CDF Method (General):** For $Y = g(X)$:
  1. Find CDF: $F_Y(y) = P(Y \le y) = P(g(X) \le y)$
  2. Express in terms of X: Solve $g(X) \le y$ for X
  3. Use CDF of X to evaluate
  4. Differentiate: $f_Y(y) = F_Y'(y)$
- **PDF Method (Monotonic g):** $\boxed{f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy}g^{-1}(y)\right|}$
- **Worked Example:** $X \sim U(0,1)$. Find PDF of $Y = X^2$.
  * CDF Method: $F_Y(y) = P(X^2 \le y) = P(X \le \sqrt{y}) = \sqrt{y}$ for $0 < y < 1$
  * Differentiate: $f_Y(y) = \frac{1}{2\sqrt{y}}$ for $0 < y < 1$
  * Verify: $\int_0^1 \frac{1}{2\sqrt{y}}dy = [\sqrt{y}]_0^1 = 1$ ✓
- **Worked Example:** $X \sim \text{Exp}(\lambda)$. Find PDF of $Y = e^X$.
  * $X = \ln Y$, so $g^{-1}(y) = \ln y$, $\frac{d}{dy}\ln y = 1/y$
  * $f_Y(y) = \lambda e^{-\lambda \ln y} \cdot \frac{1}{y} = \lambda y^{-\lambda-1}$ for $y > 1$

TWO-VARIABLE TRANSFORMATIONS (Jacobian):

- **Setup:** $(X,Y) \to (U,V)$ via $U = g_1(X,Y)$, $V = g_2(X,Y)$
- **Jacobian Method:** $\boxed{f_{UV}(u,v) = f_{XY}(x(u,v), y(u,v)) \cdot |J|}$
- **Jacobian Determinant:** $\boxed{J = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} \quad \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}}$
- **Step-by-Step Procedure:**
  1. Write transformation: $u = g_1(x,y)$, $v = g_2(x,y)$
  2. Find inverse: $x = x(u,v)$, $y = y(u,v)$
  3. Compute all 4 partial derivatives
  4. Calculate $|J|$ (ABSOLUTE VALUE!)
  5. Substitute: $f_{UV}(u,v) = f_{XY}(x(u,v), y(u,v)) \cdot |J|$
  6. Determine new support region
- **Worked Example: Polar Coordinates:**
  * $X = R\cos\Theta$, $Y = R\sin\Theta$ (inverse: $R = \sqrt{X^2+Y^2}$, $\Theta = \arctan(Y/X)$)
  * Jacobian: $J = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r\cos^2\theta + r\sin^2\theta = r$
  * $|J| = r$
  * If (X,Y) uniform on unit disk: $f_{XY}(x,y) = \frac{1}{\pi}$
  * Then $f_{R,\Theta}(r,\theta) = \frac{1}{\pi} \cdot r = \frac{r}{\pi}$ for $0 < r < 1$, $0 < \theta < 2\pi$
- **Sum and Difference:** $U = X + Y$, $V = X - Y$
  * Inverse: $X = (U+V)/2$, $Y = (U-V)/2$
  * $|J| = |-1/2| = 1/2$

! Always take ABSOLUTE VALUE of Jacobian! $|J|$ not J *Note: For marginal of U: integrate $f_{UV}(u,v)$ over v*

4.7 Order Statistics (a.k.a. Max/Min of i.i.d., $X_{(k)}$)

Terms: $X_{(1)}$ = minimum (smallest), $X_{(n)}$ = maximum (largest), $X_{(k)}$ = k-th order statistic (k-th smallest), Range = $X_{(n)} - X_{(1)}$, i.i.d. = independent and identically distributed

- **SYNONYMS:** "maximum", "minimum", "k-th smallest", "range", "largest", "smallest"
- **Definition:** Sort $X_1, ..., X_n$ to get $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$
  * $X_{(1)}$ = minimum
  * $X_{(n)}$ = maximum
  * $X_{(k)}$ = k-th smallest (order statistic)

MAXIMUM (Most Common):

- **CDF of Maximum:** If $X_1, ..., X_n$ i.i.d. with CDF F: $\boxed{F_{X_{(n)}}(x) = P(\max \le x) = P(\text{all} \le x) = [F(x)]}$
- **PDF of Maximum:** $\boxed{f_{X_{(n)}}(x) = n[F(x)]^{n-1}f(x)}$
- **Probability Max Exceeds a:** $\boxed{P(X_{(n)} > a) = 1 - [F(a)]^n}$
- **Worked Example:** $X_1, X_2, X_3 \overset{iid}{\sim} U(0,1)$. Find PDF of max.
  * $F(x) = x$ for $0 < x < 1$
  * $F_{X_{(3)}}(x) = x^3$

* $f_{X_{(3)}}(x) = 3x^2$ for $0 < x < 1$
* $E[X_{(3)}] = \int_0^1 x \cdot 3x^2 dx = 3/4$

MINIMUM:

- **CDF of Minimum:** $\boxed{F_{X_{(1)}}(x) = P(\min \le x) = 1 - P(\text{all} > x) = 1 - [1 - F(x)]^n}$
- **PDF of Minimum:** $\boxed{f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x)}$
- **Worked Example:** $X_1, ..., X_n \overset{iid}{\sim} \text{Exp}(\lambda)$. Find distribution of min.
  * $F(x) = 1 - e^{-\lambda x}$, so $1 - F(x) = e^{-\lambda x}$
  * $P(X_{(1)} > x) = [e^{-\lambda x}]^n = e^{-n\lambda x}$
  * $X_{(1)} \sim \text{Exp}(n\lambda)$ – minimum of Exp is Exp with sum of rates!

GENERAL k-th ORDER STATISTIC:

- **PDF of $X_{(k)}$:** $\boxed{f_{X_{(k)}}(x) = \dfrac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1 - F(x)]^{n-k} f(x)}$
- **Interpretation:** $k - 1$ values below x, one at x, $n - k$ above x
- **For Uniform(0,1):**
  * $X_{(k)} \sim \text{Beta}(k, n - k + 1)$
  * $E[X_{(k)}] = \frac{k}{n+1}$

USEFUL RESULTS:

- For $U(0, 1)$: $E[X_{(1)}] = \frac{1}{n+1}$, $E[X_{(n)}] = \frac{n}{n+1}$
- Range: $R = X_{(n)} - X_{(1)}$, $E[R] = \frac{n-1}{n+1}$ for $U(0, 1)$
- Median of n i.i.d. samples: $X_{((n+1)/2)}$ if n odd

Note: Order stats are useful for quality control, reliability (min lifetime), auctions (max bid)

5. MOMENT GENERATING FUNCTIONS

* Post-M2

5.1 Definition and Properties (Prof. uses $\psi(t)$ for MGF)

Terms: MGF = Moment Generating Function $M_X(t) = E[e^{tX}]$, $\psi(t)$ = professor's notation for MGF, Moment = $E[X^k]$, $M^{(k)}(0)$ = k-th derivative evaluated at $t = 0$

- **SYNONYMS:** MGF = $M_X(t)$ = $\psi(t)$ (professor's notation)
- **! Professor uses $\psi(t)$ for MGF – same thing as $M_X(t)$!**
- **Definition:** $\boxed{M_X(t) = \psi(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & \text{discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{continuous} \end{cases}}$
- **Why MGFs are Useful:**
  * Uniquely determines distribution
  * Easy to find moments
  * Product rule for sums of independent RVs
  * Can identify distribution by matching MGF

FINDING MOMENTS FROM MGF:

- **Moment Formula:** $\boxed{E[X^k] = M_X^{(k)}(0) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}}$
- **Step-by-Step:**
  1. Differentiate MGF k times with respect to t
  2. Evaluate at $t = 0$
  3. Result is $E[X^k]$
- **Worked Example:** $X \sim \text{Exp}(\lambda)$. MGF: $M(t) = \frac{\lambda}{\lambda - t}$.
  * $M'(t) = \frac{\lambda}{(\lambda-t)^2}$, so $E[X] = M'(0) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$
  * $M''(t) = \frac{2\lambda}{(\lambda-t)^3}$, so $E[X^2] = M''(0) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$
  * $\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

KEY PROPERTIES:

- **Uniqueness:** If $M_X(t) = M_Y(t)$ for all t near 0, then X and Y have same distribution
- **Linear Transform:** $\boxed{M_{aX+b}(t) = e^{bt} \cdot M_X(at)}$
- **Sum of Independent RVs:** $\boxed{M_{X+Y}(t) = M_X(t) \cdot M_Y(t)}$ (only if independent!)
- **Sum of n i.i.d.:** $\boxed{M_{S_n}(t) = [M_X(t)]^n \text{ where } S_n = X_1 +}$

COMMON MGF TABLE:

- Bernoulli(p): $1 - p + pe^t$
- Binomial(n, p): $(1 - p + pe^t)^n$
- Poisson(λ): $e^{\lambda(e^t - 1)}$
- Geometric(p): $\frac{p}{1-(1-p)e^t}$
- Uniform(a, b): $\frac{e^{tb} - e^{ta}}{t(b-a)}$
- Normal(μ, σ^2): $e^{\mu t + \sigma^2 t^2/2}$
- Exponential(λ): $\frac{\lambda}{\lambda - t}$ for $t < \lambda$
- Gamma(r, λ): $\left(\frac{\lambda}{\lambda-t}\right)^r$ for $t < \lambda$

Note: Not all distributions have MGF (e.g., Cauchy). MGF must exist in neighborhood of $t = 0$.

5.2 Using MGFs for Sums

Terms: $S_n = X_1 + \cdots + X_n$ = sum of i.i.d. RVs, Closure = sum of same-family distributions stays in that family, Product rule = $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ for independent X, Y

- **Main Technique:** Use MGF product rule to identify distribution of sum

STEP-BY-STEP PROCEDURE:

1. Let $S = X_1 + X_2 + ... + X_n$ (independent)
2. Find MGF of each: $M_{X_i}(t)$
3. Multiply: $M_S(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot ... \cdot M_{X_n}(t)$
4. For i.i.d.: $M_S(t) = [M_X(t)]^n$
5. Simplify the product
6. Match with known MGF from table $\Rightarrow$ identify distribution

IMPORTANT CLOSURE PROPERTIES:

- **Sum of Normals:**
  * $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ independent
  * $M_X(t) \cdot M_Y(t) = e^{\mu_1 t + \sigma_1^2 t^2/2} \cdot e^{\mu_2 t + \sigma_2^2 t^2/2}$
  * $= e^{(\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2)t^2/2}$
  * $\Rightarrow X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

- **Sum of Poissons:**
  - $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$ independent
  - $M_X(t) \cdot M_Y(t) = e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$
  - $\Rightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
- **Sum of Exponentials (same rate):**
  - $X_1, ..., X_n \overset{iid}{\sim} \text{Exp}(\lambda)$
  - $M_S(t) = \left(\frac{\lambda}{\lambda-t}\right)^n$
  - $\Rightarrow S = \sum X_i \sim \text{Gamma}(n, \lambda)$
- **Sum of Gammas (same rate):**
  - $X \sim \text{Gamma}(r_1, \lambda)$, $Y \sim \text{Gamma}(r_2, \lambda)$ independent
  - $\Rightarrow X + Y \sim \text{Gamma}(r_1 + r_2, \lambda)$
- **Sum of Binomials (same p):**
  - $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p)$ independent
  - $\Rightarrow X + Y \sim \text{Bin}(n_1 + n_2, p)$

WORKED EXAMPLE:

- **Problem:** $X_1, X_2, X_3 \overset{iid}{\sim} \text{Exp}(2)$. Find distribution of $S = X_1 + X_2 + X_3$.
- **Solution:**
  - MGF of Exp(2): $M_X(t) = \frac{2}{2-t}$
  - MGF of sum: $M_S(t) = \left(\frac{2}{2-t}\right)^3$
  - This matches $\text{Gamma}(3, 2)$ MGF
  - $\Rightarrow S \sim \text{Gamma}(3, 2)$
  - Check: $E[S] = 3/2$, $\text{Var}(S) = 3/4$

! MGF product only works for INDEPENDENT random variables! *Note: MGF technique is key for proving Central Limit Theorem*

6. LIMIT THEOREMS

*** Critical for Final!**

6.1 Central Limit Theorem (a.k.a. CLT, Normal Approximation)

*** Most Important!** **Terms:** CLT = Central Limit Theorem, $\bar{X}_n$ = sample mean, S_n = sum of n i.i.d. RVs, μ = mean of one observation, σ^2 = variance of one observation, $\xrightarrow{d}$ = converges in distribution

- **SYNONYMS:** CLT, "approximate", "large n", "as $n \to \infty$", "normal approximation"
- **TRIGGER WORDS:** "i.i.d.", "sample mean", "total/sum of n games", "average of n", "400 games"

CLT STATEMENT:

- If $X_1, ..., X_n$ are i.i.d. with mean μ and variance $\sigma^2 < \infty$:
- **For Sample Mean:** $\boxed{\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(\mu, \frac{\sigma^2}{n}\right)}$
- **For Sum/Total:** $\boxed{S_n = \sum_{i=1}^{n} X_i \approx N(n\mu, n\sigma^2)}$
- **Standardized Form:** $\boxed{Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)}$

STEP-BY-STEP PROCEDURE (Sample Mean):

1. **Identify:** X_i = single observation/trial
2. **Compute:** $\mu = E[X_i]$ (mean of ONE observation)
3. **Compute:** $\sigma^2 = \text{Var}(X_i)$ (variance of ONE observation)
4. **CLT:** $\bar{X} \approx N(\mu, \sigma^2/n)$
5. **Standardize:** $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
6. **Calculate:** $P(\bar{X} > c) = P(Z > \frac{c-\mu}{\sigma/\sqrt{n}}) = 1 - \Phi(\frac{c-\mu}{\sigma/\sqrt{n}})$

STEP-BY-STEP PROCEDURE (Sum/Total):

1. **Identify:** X_i = single observation/game outcome
2. **Compute:** $\mu = E[X_i]$, $\sigma^2 = \text{Var}(X_i)$
3. **CLT:** $S_n = \sum X_i \approx N(n\mu, n\sigma^2)$
4. **Standardize:** $Z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$
5. **Calculate:** Use normal table

WORKED EXAMPLE: 400 games. Win \$3 (prob 1/4), lose \$1 (prob 1/4), else \$0 (prob 1/2).

- **Step 1:** X_i = profit from one game
- **Step 2:** $E[X] = 3(1/4) + (-1)(1/4) + 0(1/2) = 3/4 - 1/4 = 1/2$
- **Step 3:** $E[X^2] = 9(1/4) + 1(1/4) + 0 = 10/4 = 5/2$
- **Step 4:** $\text{Var}(X) = 5/2 - 1/4 = 10/4 - 1/4 = 9/4$, so $\sigma = 3/2$
- **Step 5:** Total $S_{400} \approx N(400 \cdot 1/2, 400 \cdot 9/4) = N(200, 900)$
- **Step 6:** $P(S_{400} > 240) = P(Z > \frac{240-200}{30}) = P(Z > 4/3) = 1 - \Phi(1.33) \approx 0.092$

WHEN TO USE CLT:

- $n \geq 30$ (rule of thumb)
- For skewed distributions, may need larger n
- Works for ANY distribution with finite variance

! For discrete RVs, apply continuity correction (see 6.2)!

6.2 Normal Approximations

Terms: Continuity correction = adjust by ± 0.5 when approximating discrete with continuous, Normal approximation = use $N(\mu, \sigma^2)$ to approximate discrete distribution, Rule of thumb = $np \geq 10$ and $n(1-p) \geq 10$ for Binomial

BINOMIAL APPROXIMATION:

- **When to Use:** $X \sim \text{Binomial}(n, p)$ with $np \geq 10$ AND $n(1-p) \geq 10$
- **Approximation:** $\boxed{X \approx N(np, np(1-p))}$
- **Worked Example:** $X \sim \text{Bin}(100, 0.3)$. Find $P(X > 35)$.

* Check: $np = 30 \geq 10$ ✓, $n(1-p) = 70 \geq 10$ ✓
* $\mu = np = 30$, $\sigma^2 = np(1-p) = 21$, $\sigma = \sqrt{21} \approx 4.58$
* With continuity correction: $P(X > 35) = P(X \geq 36) \approx P(Y > 35.5)$
* $Z = (35.5 - 30)/4.58 = 1.20$
* $P(X > 35) \approx 1 - \Phi(1.20) = 0.115$

POISSON APPROXIMATION:

- **When to Use:** $X \sim \text{Poisson}(\lambda)$ with $\lambda \geq 30$
- **Approximation:** $\boxed{X \approx N(\lambda, \lambda)}$
  * Mean = Variance = λ for Poisson

CONTINUITY CORRECTION (Critical for Discrete!):

- **Why:** Discrete integer $\to$ continuous normal; adjust by ± 0.5
- **Rules (let Y = normal approximation):**

Discrete Prob	Normal Approx
$P(X = k)$	$P(k - 0.5 < Y < k + 0.5)$
$P(X \leq k)$	$P(Y < k + 0.5)$
$P(X < k)$	$P(Y < k - 0.5)$
$P(X \geq k)$	$P(Y > k - 0.5)$
$P(X > k)$	$P(Y > k + 0.5)$
$P(a \leq X \leq b)$	$P(a - 0.5 < Y < b + 0.5)$

- **Memory Aid:**
  * "$\leq k$" includes k, so go to $k + 0.5$
  * "$< k$" excludes k, so stop at $k - 0.5$
  * "$\geq k$" includes k, so start at $k - 0.5$
  * "$> k$" excludes k, so start at $k + 0.5$
- **Worked Example:** $X \sim \text{Bin}(50, 0.4)$. Find $P(X = 20)$.
  * $\mu = 20$, $\sigma = \sqrt{50 \times 0.4 \times 0.6} = \sqrt{12} \approx 3.46$
  * $P(X = 20) \approx P(19.5 < Y < 20.5)$
  * $= \Phi(\frac{20.5-20}{3.46}) - \Phi(\frac{19.5-20}{3.46}) = \Phi(0.14) - \Phi(-0.14)$
  * $= 0.556 - 0.444 = 0.112$

! Forgetting continuity correction is a common exam mistake!

6.3 Law of Large Numbers (LLN)

Terms: LLN = Law of Large Numbers, $\xrightarrow{P}$ = converges in probability, $\xrightarrow{a.s.}$ = almost sure convergence, Weak LLN = convergence in probability, Strong LLN = almost sure convergence

- **Statement:** As $n \to \infty$, sample mean $\to$ true mean
- **Weak LLN:** $\boxed{\bar{X}_n \xrightarrow{P} \mu}$ (convergence in probability)
  * For any $\epsilon > 0$: $\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$
- **Strong LLN:** $\boxed{\bar{X}_n \xrightarrow{a.s.} \mu}$ (almost sure convergence)
  * $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$
- **Intuition:**
  * Flip fair coin many times: proportion of heads $\to 0.5$
  * Casino: average winnings per game $\to$ expected value
  * Sample mean is consistent estimator of population mean
- **LLN vs CLT:**
  * **LLN:** WHERE does $\bar{X}_n$ converge? (Answer: to μ)
  * **CLT:** HOW is $\bar{X}_n$ distributed? (Answer: approximately normal)
- **Worked Example:** Roll die 1000 times. By LLN, average roll $\approx E[X] = 3.5$
- *Note: LLN justifies using relative frequency as probability estimate*

6.4 Confidence Intervals

Terms: CI = Confidence Interval, $(1-\alpha)$ = confidence level (e.g., 95%), α = significance level, $z_{\alpha/2}$ = critical value from standard normal, SE = standard error = $\sigma/\sqrt{n}$, ME = margin of error

- **Definition:** Interval estimate with specified confidence level $(1-\alpha)$
- **CI for Mean (known σ):** $\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/}\right.$
  $\boxed{\text{or } \bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$

CRITICAL VALUES:

Confidence	α	$z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.576

STEP-BY-STEP PROCEDURE:

1. Identify confidence level $(1-\alpha)$ and find $z_{\alpha/2}$
2. Compute sample mean $\bar{X}$
3. Compute standard error: $SE = \sigma/\sqrt{n}$
4. Margin of error: $ME = z_{\alpha/2} \cdot SE$
5. CI: $(\bar{X} - ME, \bar{X} + ME)$

WORKED EXAMPLE: $n = 100$, $\bar{X} = 52$, $\sigma = 10$ (known). Find 95% CI.

- $z_{0.025} = 1.96$
- $SE = 10/\sqrt{100} = 1$
- $ME = 1.96 \times 1 = 1.96$
- 95% CI: $(52 - 1.96, 52 + 1.96) = (50.04, 53.96)$

INTERPRETATION (Critical!):

- **CORRECT:** "95% of intervals constructed this way contain μ"
- **! WRONG:** "95% probability that μ is in this interval"
- The parameter μ is fixed; the interval is random

SAMPLE SIZE DETERMINATION:

- **Given desired margin of error E:** $\boxed{n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2}$
- **Example:** Want 95% CI with margin ± 2, $\sigma = 10$.
  * $n \geq (1.96 \times 10/2)^2 = (9.8)^2 = 96.04$
  * Need $n \geq 97$

CI PROPERTIES:

- Larger $n \Rightarrow$ narrower CI (more precision)
- Higher confidence $\Rightarrow$ wider CI (more certainty requires wider net)
- Width $= 2 \times z_{\alpha/2} \times \sigma/\sqrt{n}$

Note: CI width halves when n quadruples (due to $\sqrt{n}$)

7. SPECIAL TOPICS & APPLICATIONS

*** Post-M2**

7.1 Conditional Expectation (a.k.a. $E[X|Y]$, Total Expectation)

*** Conceptual!** **Terms:** $E[X|Y = y]$ = conditional expectation (a number), $E[X|Y]$ = conditional expectation (a random variable, function of Y), Tower property = $E[X]$ = $E[E[X|Y]]$, Total expectation = law of iterated expectations

- **SYNONYMS:** "$E[X|Y]$", "average given", "expected value given"
- **TRIGGER WORDS:** "$E[X|Y = y]$", "break down by cases", "tower property", "given"

DEFINITION:

- $E[X|Y = y]$ **(a number):**
  - Discrete: $E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y)$
  - Continuous: $E[X|Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\,dx$
- $E[X|Y]$ **(a random variable):**
  - $E[X|Y]$ is a function of Y, so it's a random variable!
  - Substitute Y for y in $E[X|Y = y]$

LAW OF TOTAL EXPECTATION (Tower Property):

- $\boxed{E[X] = E[E[X|Y]]}$
- **Discrete version:** $E[X] = \sum_y E[X|Y = y] \cdot P(Y = y)$
- **Continuous version:** $E[X] = \int E[X|Y = y] \cdot f_Y(y)\,dy$
- **Intuition:** Average of conditional averages = overall average

STEP-BY-STEP: Computing $E[X]$ via Conditioning:

1. Choose what to condition on (often a natural "first stage")
2. Compute $E[X|Y = y]$ for each value of Y
3. Apply: $E[X] = \sum_y E[X|Y = y] \cdot P(Y = y)$ or integral

WORKED EXAMPLE: Roll die. If outcome is Y, flip Y coins. Let X = # heads.

- $X|Y = y \sim \text{Bin}(y, 1/2)$, so $E[X|Y = y] = y/2$
- $E[X|Y] = Y/2$ (random variable)
- $E[X] = E[Y/2] = E[Y]/2 = 3.5/2 = 1.75$

LAW OF TOTAL VARIANCE:

- $\boxed{\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])}$
- **Interpretation:**
  - $E[\text{Var}(X|Y)]$ = average within-group variance
  - $\text{Var}(E[X|Y])$ = variance of group means

KEY PROPERTIES:

- Linearity: $E[aX + bZ|Y] = aE[X|Y] + bE[Z|Y]$
- Taking out known: $E[h(Y)\cdot X|Y] = h(Y)\cdot E[X|Y]$
- If X, Y independent: $E[X|Y] = E[X]$
- Tower: $E[E[X|Y, Z]|Y] = E[X|Y]$

Note: $E[X|Y]$ is the best predictor of X given Y (minimizes MSE)

7.2 Bayesian Statistics (a.k.a. Prior/Posterior, Conjugate Priors)

*** Professor's Favorite!** **Terms:** Prior $\pi(\theta)$ = belief before data, Likelihood $L(x|\theta)$ = prob of data given parameter, Posterior $\pi(\theta|x)$ = updated belief after data, Conjugate prior = posterior is same family as prior, MAP = Maximum A Posteriori (mode of posterior)

- **SYNONYMS:** "prior", "posterior", "update belief", "given evidence", "conjugate"
- **TRIGGER WORDS:** "defective rate", "unknown parameter", "given data", "Monty Hall"

BAYESIAN FRAMEWORK:

- $\boxed{\text{Posterior} \propto \text{Likelihood} \times \text{Prior}}$
- $\boxed{\pi(\theta|x) = \dfrac{L(x|\theta)\pi(\theta)}{\int L(x|\theta)\pi(\theta)d\theta}}$
- **Components:**
  - $\pi(\theta)$ = Prior: belief about θ BEFORE seeing data
  - $L(x|\theta) = P(\text{data}|\theta)$ = Likelihood: prob of data given parameter
  - $\pi(\theta|x)$ = Posterior: updated belief AFTER seeing data
  - $d\theta$ = "integrate over all possible values of θ" (the denominator)
  - $\int L(x|\theta)\pi(\theta)d\theta$ = normalizing constant (makes posterior integrate to 1)

WHAT IS A KERNEL? (Critical for Recognition!)

- **Kernel:** The part of a PDF that depends on the variable (ignoring constants)
- **How to use:** Match the θ-dependent part to a known distribution
- **Common Kernels to Recognize:**
  - $\theta^{a-1}(1-\theta)^{b-1} \Rightarrow \text{Beta}(a, b)$ (for $0 < \theta < 1$)
  - $\theta^{a-1}e^{-b\theta} \Rightarrow \text{Gamma}(a, b)$ (for $\theta > 0$)
  - $e^{-(\theta-\mu)^2/(2\sigma^2)} \Rightarrow \text{Normal}(\mu, \sigma^2)$

  * $e^{-\lambda\theta} \Rightarrow$ Exponential(λ) (for $\theta > 0$)
- **Example:** If posterior $\propto \theta^4(1-\theta)^7$, recognize as Beta$(5, 8)$

HOW TO NORMALIZE (when kernel not recognized):

1. Compute $c = \int L(x|\theta)\pi(\theta)d\theta$ over support of θ
2. Posterior PDF $= \frac{1}{c} \cdot L(x|\theta)\pi(\theta)$
3. **Shortcut:** If you recognize the kernel, the normalizing constant is known!
   - Beta(a, b): constant $= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = B(a, b)$
   - Gamma(a, b): constant $= \frac{\Gamma(a)}{b^a}$

STEP-BY-STEP: Bayesian Update (Continuous Prior):

1. Specify prior distribution $\pi(\theta)$
2. Write likelihood function $L(x|\theta)$
3. Multiply: $\pi(\theta|x) \propto L(x|\theta) \cdot \pi(\theta)$
4. Collect all θ-dependent terms (the kernel)
5. Recognize the kernel $\Rightarrow$ identify distribution family and parameters
6. Or normalize: divide by $\int L(x|\theta)\pi(\theta)d\theta$

CONJUGATE PRIORS (Posterior same family as prior):
Key: n = # observations, x = # successes, $\sum x_i$ = sum of data

- **Binomial**(n, θ) + **Beta**(α, β)**:** Posterior = Beta$(\alpha + x, \beta + n - x)$
  * α += successes, β += failures
- **Poisson**(θ) + **Gamma**(α, β)**:** Posterior = Gamma$(\alpha + \sum x_i, \beta + n)$
  * α += total count, β += # observations
- **Exp**(θ) + **Gamma**(α, β)**:** Posterior = Gamma$(\alpha + n, \beta + \sum x_i)$
  * α += # observations, β += sum of data
- **Normal**(θ, σ^2) + **Normal**(μ_0, τ^2)**:** Posterior = Normal (precision-weighted avg)
  * Posterior mean $= \frac{\tau^2\bar{x} + (\sigma^2/n)\mu_0}{\tau^2 + \sigma^2/n}$

FULL WORKED EXAMPLE: Beta-Binomial Update:
Problem: Estimate defect rate θ. Prior belief: $\theta \sim$ Beta$(2, 3)$. Observe 7 defectives in 20 items.

1. **Prior:** $\theta \sim$ Beta$(\alpha = 2, \beta = 3)$
   - Prior mean: $E[\theta] = \frac{\alpha}{\alpha+\beta} = \frac{2}{5} = 0.4$
   - Prior variance: $\text{Var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{6}{25 \cdot 6} = 0.04$
2. **Likelihood:** $X|\theta \sim$ Binomial$(n = 20, \theta)$, observed $x = 7$
$$L(\theta) = \binom{20}{7}\theta^7(1-\theta)^{13} \propto \theta^7(1-\theta)^{13}$$
3. **Posterior $\propto$ Likelihood $\times$ Prior:**
$$\pi(\theta|x) \propto \theta^7(1-\theta)^{13} \cdot \theta^{2-1}(1-\theta)^{3-1} = \theta^8(1-\theta)^{15}$$
4. **Recognize kernel:** $\theta^8(1-\theta)^{15} = \theta^{9-1}(1-\theta)^{16-1} \Rightarrow$ Beta$(9, 16)$
5. **Posterior:** $\theta|x \sim$ Beta$(\alpha' = 2 + 7 = 9, \beta' = 3 + 13 = 16)$
   - Posterior mean: $E[\theta|x] = \frac{9}{9+16} = \frac{9}{25} = 0.36$
   - Posterior variance: $\text{Var}(\theta|x) = \frac{9 \cdot 16}{25^2 \cdot 26} = \frac{144}{16250} \approx 0.0089$
6. **Interpretation:** Data (35% defective) pulled estimate down from 0.4 to 0.36. Variance decreased (more certain after seeing data).

DISCRETE PRIOR (Finite hypotheses):

- When $\theta \in \{\theta_1, \theta_2, ..., \theta_k\}$
- Use Bayes' theorem: $P(\theta_i|x) = \frac{P(x|\theta_i)P(\theta_i)}{\sum_j P(x|\theta_j)P(\theta_j)}$
- Same as Section 1.3 but with parameter interpretation

POSTERIOR MEAN AND MAP:

- Posterior Mean: $E[\theta|x] = \int \theta \cdot \pi(\theta|x)d\theta$
- MAP (Maximum A Posteriori): Mode of posterior
- For Beta(α, β): Mean $= \frac{\alpha}{\alpha+\beta}$, Mode $= \frac{\alpha-1}{\alpha+\beta-2}$

Note: Conjugate priors make computation easy – posterior is same family as prior!

7.3 Lognormal Distribution (a.k.a. $\ln X \sim N(\mu, \sigma^2)$)

*** Finance Applications!** **Terms:** Lognormal = $Y = e^X$ where $X \sim N(\mu, \sigma^2)$, μ = mean of $\ln Y$, σ^2 = variance of $\ln Y$, Log return = $\ln(S_T/S_0)$, Geometric Brownian Motion = stock price model

- **SYNONYMS:** Lognormal, "log X is normal", "e^X where $X \sim N$", "stock price model"
- **TRIGGER WORDS:** "stock price", "$S = S_0e^Z$", "log returns", "always positive"

DEFINITION AND PROPERTIES:

- **Definition:** If $X \sim N(\mu, \sigma^2)$, then $Y = e^X$ is Lognormal
- **Equivalently:** $\ln Y \sim N(\mu, \sigma^2)$
- **Key Parameters:**
  * Mean: $\boxed{E[Y] = e^{\mu+\sigma^2/2}}$
  * Variance: $\text{Var}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$
  * Median: e^μ (note: median $\neq$ mean due to skewness)
  * Mode: $e^{\mu-\sigma^2}$
- **Properties:**
  * Always positive: $Y > 0$
  * Right-skewed
  * Product of lognormals is lognormal

CRITICAL FORMULA:

- $\boxed{E[e^X] = e^{\mu+\sigma^2/2} \text{ when } X \sim N(\mu, \sigma^2)}$
- This is THE key formula for lognormal problems!

STOCK PRICE MODEL:

- **Model:** $S_T = S_0e^Z$ where $Z \sim N(\mu, \sigma^2)$
- **Risk-Neutral:** Often $Z \sim N((r - \sigma^2/2)T, \sigma^2T)$ where r = risk-free rate
- **Key Result:** $E[e^{-rT}S_T] = S_0$ (discounted expected price = current price)

GENERALIZED LOGNORMAL STOCK PRICE WALKTHROUGH:
Given: $S = S_0 e^Z$ *where* $Z \sim N(r - \sigma^2/2, \sigma^2)$, *with* S_0, r, σ *given.*

Part A: Computing $E[e^{-r}S]$ (Discounted Expected Value):

1. **Recognize:** MGF of standard normal is $E[e^{tW}] = e^{t^2/2}$ for $W \sim N(0,1)$
2. **Rewrite Z:** $Z = (r - \sigma^2/2) + \sigma \cdot N(0,1)$
3. **Compute:**

$$\begin{aligned} E[e^{-r}S] &= e^{-r}S_0E[e^Z] = e^{-r}S_0E[e^{(r-\sigma^2/2)+\sigma N(0,1)}] \\ &= e^{-r}S_0e^{(r-\sigma^2/2)}E[e^{\sigma N(0,1)}] \\ &= e^{-r}S_0e^{(r-\sigma^2/2)}e^{\sigma^2/2} = S_0 \end{aligned}$$

4. **Key insight:** Discounted expected price equals current price (risk-neutral pricing)

Part B: Computing $P(S > K)$ (Probability Exceeds Threshold):

1. **Setup:** $P(S > K) = P(S_0e^Z > K)$
2. **Take log:** $P(Z > \ln(K/S_0))$
3. **Standardize:** Since $Z \sim N(r - \sigma^2/2, \sigma^2)$:

$$P\left(\frac{Z-(r-\sigma^2/2)}{\sigma} > \frac{\ln(K/S_0)-(r-\sigma^2/2)}{\sigma}\right)$$

4. **Use symmetry:** Let $d = \frac{\ln(K/S_0)-(r-\sigma^2/2)}{\sigma}$

$$P(N(0,1) > d) = 1 - \Phi(d) = \Phi(-d)$$

5. **Final formula:** $\boxed{P(S > K) = \Phi\left(\frac{(r-\sigma^2/2)-\ln(K/S_0)}{\sigma}\right) = \Phi\left(\frac{\ln(S_0/K)+r-\sigma^2/2}{\sigma}\right)}$

Quick Reference for $S = S_0e^Z$, $Z \sim N(r - \sigma^2/2, \sigma^2)$:

- $E[S] = S_0e^r$ (expected price grows at rate r)
- $E[e^{-r}S] = S_0$ (discounted expected = current price)
- $P(S > K) = \Phi\left(\frac{\ln(S_0/K)+r-\sigma^2/2}{\sigma}\right)$

STEP-BY-STEP: Finding $P(S > K)$ (General):

1. Write $S = S_0e^Z$ where $Z \sim N(\mu, \sigma^2)$
2. $P(S > K) = P(S_0e^Z > K) = P(e^Z > K/S_0)$
3. $= P(Z > \ln(K/S_0))$
4. $= 1 - \Phi\left(\frac{\ln(K/S_0)-\mu}{\sigma}\right)$

WORKED EXAMPLE:

- $S_0 = 100$, $Z \sim N(0.03, 0.04)$ (so $\mu = 0.03$, $\sigma = 0.2$)
- **Find $E[S]$:**
  * $E[S] = S_0 \cdot E[e^Z] = 100 \cdot e^{0.03+0.04/2} = 100 \cdot e^{0.05} \approx 105.13$
- **Find $P(S > 110)$:**
  * $P(Z > \ln(110/100)) = P(Z > 0.0953)$
  * $= P\left(\frac{Z-0.03}{0.2} > \frac{0.0953-0.03}{0.2}\right) = P(W > 0.327)$
  * $= 1 - \Phi(0.327) \approx 0.372$

[$] Lognormal is foundation of Black-Scholes option pricing *Note: Log returns* $\ln(S_T/S_0) = Z$ *are normal; prices* S_T *are lognormal*

7.4 Additional Important Concepts

Terms: I_A = indicator of event A (1 if A occurs, 0 otherwise), Markov inequality = $P(X \geq a) \leq E[X]/a$, Chebyshev inequality = $P(|X - \mu| \geq k\sigma) \leq 1/k^2$, Jensen inequality = $E[g(X)] \geq g(E[X])$ for convex g

INDICATOR RANDOM VARIABLES:

- **Definition:** $I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$
- **Key Properties:**
  * $E[I_A] = P(A)$
  * $I_A^2 = I_A$ (since 0 and 1 squared are themselves)
  * $\text{Var}(I_A) = P(A)(1 - P(A))$
  * $I_A \cdot I_B = I_{A\cap B}$
- **Useful for Counting:** If X = count of events $A_1, ..., A_n$:
  * $X = \sum_{i=1}^n I_{A_i}$
  * $E[X] = \sum_{i=1}^n E[I_{A_i}] = \sum_{i=1}^n P(A_i)$
- **Example:** Expected number of matches in a random derangement

IMPORTANT INEQUALITIES:

- **Markov's Inequality:** For $X \geq 0$ and $a > 0$: $\boxed{P(X \geq a) \leq \frac{E[X]}{a}}$
- **Chebyshev's Inequality:** For any $k > 0$: $\boxed{P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}}$
  * Equivalently: $P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$
  * At least 75% of data within 2 std devs
  * At least 89% within 3 std devs
- **Jensen's Inequality:** If g is convex: $\boxed{E[g(X)] \geq g(E[X])}$
  * For concave g: $E[g(X)] \leq g(E[X])$
  * Example: $E[X^2] \geq (E[X])^2$ (since x^2 is convex)
  * Example: $E[\ln X] \leq \ln(E[X])$ (since ln is concave)

PROBABILITY INTEGRAL TRANSFORM:

- **Theorem:** If X has continuous CDF F, then $F(X) \sim U(0,1)$
- **Inverse:** If $U \sim U(0,1)$, then $F^{-1}(U)$ has CDF F
- **Application:** Generate random variables from any distribution using uniform

8. PRACTICE PROBLEM COMPENDIUM

8.1 Bayesian Problems

*** High Frequency! Problem [HW6-1]: Monty Hall (Sober vs Dizzy)**
Contestant picks door A. Monty opens door B showing goat.

Solution:

1. **Problem Type:** Bayesian update with different likelihoods
2. **Required Concepts:** Bayes' theorem, conditional probability
3. **Sober Monty:**
   - Prior: $P(H_A) = P(H_B) = P(H_C) = 1/3$
   - Likelihood: $P(\text{open B}|H_A) = 1/2$, $P(\text{open B}|H_B) = 0$, $P(\text{open B}|H_C) = 1$
   - Posterior: $P(H_A|\text{data}) = 1/3$, $P(H_C|\text{data}) = 2/3$
   - **Strategy: Switch! (doubles probability)**
4. **Dizzy Monty:**
   - Likelihood: $P(\text{open B}|H_A) = 1/2$, $P(\text{open B}|H_B) = 1/2$, $P(\text{open B}|H_C) = 1/2$
   - Posterior: All equal at 1/3
   - **Strategy: Doesn't matter!**
5. **Key Insight:** Knowledge affects likelihood function

8.2 CLT Applications

*** Guaranteed on Final! Problem [Practice Final-1]: Coin Game with 400 Plays**
Win \$3 if HH, lose \$1 if TT, else \$0. Play 400 times.

Solution:

1. **Problem Type:** CLT with discrete outcomes
2. **Step 1:** Find distribution of single game
   - $P(X = 3) = 1/4$ (HH)
   - $P(X = -1) = 1/4$ (TT)
   - $P(X = 0) = 1/2$ (HT or TH)
3. **Step 2:** Calculate μ and σ^2
   - $E[X] = 3(1/4) + (-1)(1/4) + 0(1/2) = 0.5$
   - $E[X^2] = 9(1/4) + 1(1/4) + 0 = 2.5$
   - $\text{Var}(X) = 2.5 - 0.25 = 2.25$, so $\sigma = 1.5$
4. **Step 3:** Apply CLT for $n = 400$
   - Total: $S_{400} \approx N(400 \cdot 0.5, 400 \cdot 2.25) = N(200, 900)$
   - $P(S_{400} \geq 240) = P(Z \geq \frac{240-200}{30}) = P(Z \geq 1.33) \approx 0.092$
5. **Key Insight:** Use continuity correction: $P(S \geq 240) \approx P(S > 239.5)$

8.3 Bivariate Normal

*** Complex but Common! Problem [HW5-1]: Joint Normal with Correlation**
$X \sim N(1, 2)$, $Y \sim N(-2, 3)$, $\rho = -2/3$. Find $P(X + Y > 0)$

Solution:

1. **Problem Type:** Linear combination of bivariate normal
2. **Key Property:** $X + Y$ is normal
3. **Parameters of $Z = X + Y$:**
   - $\mu_Z = \mu_X + \mu_Y = 1 + (-2) = -1$
   - $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = 2 + 3 + 2(-2/3)\sqrt{6} = 5 - \frac{4\sqrt{6}}{3}$
4. **Standardize and compute:**
   - $P(Z > 0) = P\left(\frac{Z+1}{\sigma_Z} > \frac{1}{\sigma_Z}\right) = 1 - \Phi(0.759) \approx 0.224$
5. **Key Insight:** Always check if linear combination, use properties of bivariate normal

8.4 Joint Distributions

Problem [HW4-1]: Joint PDF Analysis
$f(x, y) = c(x^2 + xy)$ on $[0, 1] \times [0, 1]$

Solution:

1. **Find constant c:**
   - $\int_0^1 \int_0^1 (x^2 + xy)dxdy = \int_0^1 [\frac{x^3}{3} + \frac{x^2 y}{2}]_0^1 dy = \int_0^1 (\frac{1}{3} + \frac{y}{2})dy = \frac{7}{12}$
   - Therefore $c = \frac{12}{7}$
2. **Marginal of X:**
   - $f_X(x) = \int_0^1 \frac{12}{7}(x^2 + xy)dy = \frac{12}{7}x^2 + \frac{6x}{7}$
3. **Check independence:**
   - Need $f(x, y) = f_X(x) \cdot f_Y(y)$ for all (x, y)
   - Since $f(x, y)$ has xy term, NOT independent
4. **Key Insight:** Cross-product terms indicate dependence

8.5 Lognormal Distribution

[\$] Finance Focus! Problem [Practice Final-4]: Stock Price Model
$S = S_0 e^Z$ where $Z \sim N((r - \sigma^2/2), \sigma^2)$, $S_0 = 100$, $r = 0.05$, $\sigma = 0.2$

Solution:

1. **Problem Type:** Lognormal application
2. **Part (a):** Find $E[e^{-r}S] = E[S_0 e^{Z-r}]$
   - $Z - r \sim N(-\sigma^2/2, \sigma^2)$
   - $E[e^{Z-r}] = \exp(-\sigma^2/2 + \sigma^2/2) = 1$
   - Therefore $E[e^{-r}S] = S_0 = 100$
3. **Part (b):** Find $P(S > 100)$
   - $P(S > 100) = P(e^Z > 1) = P(Z > 0)$
   - $Z \sim N(-0.02, 0.04)$
   - $P(Z > 0) = P\left(\frac{Z+0.02}{0.2} > 0.1\right) = 1 - \Phi(0.1) \approx 0.46$
4. **Key Insight:** Stock prices lognormal $\Rightarrow$ log returns normal

8.6 Exponential Memoryless

Problem [Practice Final-3]: Average of Exponentials
$X_1, ..., X_{100}$ iid Exp(1/3). Find $P(\bar{X}/(\bar{X}+3) < 0.5)$

Solution:

1. **Problem Type:** CLT for exponentials
2. **Setup:** $E[X_i] = 3$, $\text{Var}(X_i) = 9$

3. **Apply CLT:** $\bar{X} \approx N(3, 9/100) = N(3, 0.09)$
4. **Transform inequality:**
   - $\frac{\bar{X}}{\bar{X}+3} < 0.5 \Rightarrow \bar{X} < 0.5(\bar{X}+3) \Rightarrow \bar{X} < 3$
5. **Calculate:** $P(\bar{X} < 3) = 0.5$ (by symmetry of normal)
6. **Key Insight:** Transform inequality first, then apply CLT

8.7 Order Statistics

Problem: Max and Min of Uniform(0,1)
$X_1, ..., X_n$ iid Uniform(0,1). Find distribution of max and min.

Solution:

1. **Maximum $X_{(n)}$:**
   - $F_{max}(x) = P(\text{all} \le x) = x^n$
   - $f_{max}(x) = nx^{n-1}$ for $0 < x < 1$
   - $E[X_{(n)}] = \frac{n}{n+1}$
2. **Minimum $X_{(1)}$:**
   - $F_{min}(x) = 1 - P(\text{all} > x) = 1 - (1-x)^n$
   - $f_{min}(x) = n(1-x)^{n-1}$ for $0 < x < 1$
   - $E[X_{(1)}] = \frac{1}{n+1}$
3. **Key Insight:** Use complement for min, direct for max

8.8 Conjugate Priors

*** Bayesian Favorite! Problem [Practice Final-5]: Beta-Binomial Update**
Prior: $\theta \in \{1/2, 3/4\}$ equally likely. Data: 0 defects in 10 items.

Solution:

1. **Problem Type:** Discrete prior Bayesian update
2. **Likelihoods:**
   - $P(0 \text{ defects}|\theta = 1/2) = (1/2)^{10} = 1/1024$
   - $P(0 \text{ defects}|\theta = 3/4) = (1/4)^{10} = 1/1048576$
3. **Posterior:**
   - $P(\theta = 1/2|\text{data}) \propto (1/2) \cdot 1/1024 = 1/2048$
   - $P(\theta = 3/4|\text{data}) \propto (1/2) \cdot 1/1048576 \approx 0$
   - After normalization: $P(\theta = 1/2|\text{data}) \approx 0.999$
4. **Key Insight:** Extreme data strongly favors lower defect rate

8.9 Conditional Expectation

Problem: Breaking Sticks
Break at $X \sim U(0, \ell)$, then break smaller piece at $Y|X \sim U(0, X)$

Solution:

1. **Joint density:** $f(x, y) = \frac{1}{\ell} \cdot \frac{1}{x} = \frac{1}{\ell x}$ for $0 < y < x < \ell$
2. **Marginal of Y:** $f_Y(y) = \int_y^\ell \frac{1}{\ell x} dx = \frac{1}{\ell} \ln(\ell/y)$
3. **Conditional expectation:** $E[Y|X] = X/2$
4. **Total expectation:** $E[Y] = E[E[Y|X]] = E[X/2] = \ell/4$
5. **Total variance:** Use $\text{Var}(Y) = E[\text{Var}(Y|X)] + \text{Var}(E[Y|X])$
6. **Key Insight:** Hierarchical structure leads to law of total expectation

8.10 Hypothesis Testing & Confidence Intervals

Problem: Test Average with CLT
Sample mean $\bar{X} = 52$ from $n = 100$, known $\sigma = 10$. Test $H_0 : \mu = 50$

Solution:

1. **Test statistic:** $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{52-50}{10/10} = 2$
2. **P-value:** $P(|Z| > 2) = 2(1 - \Phi(2)) = 0.0455$
3. **95% CI:** $\bar{X} \pm 1.96 \cdot \sigma/\sqrt{n} = 52 \pm 1.96 = [50.04, 53.96]$
4. **Decision:** Reject H_0 at 5% level (barely)
5. **Key Insight:** CI excludes 50, consistent with rejection

9. MULTI-STEP PROBLEM TEMPLATES

*** Critical!**

Template A: Gaussian Vector Problems

When you see: "Gaussian vector", "independent components", "MVN", "jointly normal"

Step-by-Step Procedure:

1. **Recognize terminology:** "Gaussian" = Normal, "Gaussian vector" = MVN
2. **For independent components:** Set $\rho = 0$ (equivalently $\text{Cov}(Y_1, Y_2) = 0$)
3. **Compute covariance:** Use $\text{Cov}(aX + bY, cX + dY) = ac\text{Var}(X) + bd\text{Var}(Y) + (ad + bc)\text{Cov}(X, Y)$
4. **For i.i.d. $N(0, 1)$:** $\text{Cov}(Y_1, Y_2)$ = sum of products of coefficients
5. **Solve for parameter:** Set covariance = 0
6. **Write joint density:** Product of marginal densities (since independent)

Key Formula: For $Y_1 = aX_1 + X_2$, $Y_2 = X_1 + bX_2$ where $X_1, X_2 \overset{iid}{\sim} N(0, 1)$:

- $\text{Cov}(Y_1, Y_2) = a \cdot 1 \cdot 1 + 1 \cdot b \cdot 1 = a + b$ (since $\text{Var}(X_i) = 1$, $\text{Cov}(X_1, X_2) = 0$)
- Independence requires: $a + b = 0 \Rightarrow b = -a$
- Marginals: $Y_1 \sim N(0, a^2+1)$, $Y_2 \sim N(0, 1+a^2)$
- Joint density: $f(y_1, y_2) = \frac{1}{2\pi(a^2+1)} \exp\left(-\frac{y_1^2+y_2^2}{2(a^2+1)}\right)$

Verification: Check that joint = product of marginals!

Template B: CLT Game/Coin Problems

When you see: "400 games", "total winnings", "approximate", "n trials"

Complete Step-by-Step:

1. **Define RV:** X_i = outcome/profit from single game i
2. **List PMF:** Create table of all outcomes with probabilities
3. **Compute $E[X]$:** $E[X_i] = \sum_x x \cdot P(X = x)$
4. **Compute $E[X^2]$:** $E[X_i^2] = \sum_x x^2 \cdot P(X = x)$
5. **Compute $\text{Var}(X)$:** $\text{Var}(X_i) = E[X^2] - (E[X])^2$
6. **Apply CLT for sum:** $S_n = \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2)$
7. **Standardize:** $Z = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$
8. **Calculate probability:** Use normal table
9. **Apply continuity correction if needed:** For discrete S_n

Full Worked Example: Win \$3 if HH, lose \$1 if TT, else \$0. Play 400 times.

- PMF: $P(X = 3) = 1/4$, $P(X = -1) = 1/4$, $P(X = 0) = 1/2$
- $E[X] = 3(1/4) + (-1)(1/4) + 0(1/2) = 3/4 - 1/4 = 1/2$
- $E[X^2] = 9(1/4) + 1(1/4) + 0 = 10/4$
- $\text{Var}(X) = 10/4 - (1/2)^2 = 10/4 - 1/4 = 9/4$, $\sigma = 3/2$
- $S_{400} \approx N(400 \cdot 1/2, 400 \cdot 9/4) = N(200, 900)$
- $P(S > 240) = P(Z > \frac{240-200}{30}) = P(Z > 1.33) \approx 0.092$

Verification: Does $n\mu = 200$ make sense? (Expected to win \$200 in 400 games at \$0.50/game) ✓

Template C: Exponential + CLT

When you see: "i.i.d. Exp", "mean θ", "average"

! If "mean $\theta = 3$" then $\lambda = 1/3$ NOT 3!

Steps:

1. **Parameters:** $E[X_i] = 1/\lambda$, $\text{Var}(X_i) = 1/\lambda^2$
2. **For $\bar{X}$:** $E[\bar{X}] = 1/\lambda$, $\text{Var}(\bar{X}) = 1/(n\lambda^2)$
3. **CLT:** $\bar{X} \approx N(1/\lambda, 1/(n\lambda^2))$
4. **Transform inequality first:** e.g., $\bar{X}/(\bar{X} + 3) < 0.5 \Rightarrow \bar{X} < 3$

Template D: Lognormal Stock Price

When you see: "$S = S_0 e^Z$", "$Z \sim N(\mu, \sigma^2)$"

Key Formulas:

- $E[e^Z] = e^{\mu + \sigma^2/2}$ when $Z \sim N(\mu, \sigma^2)$
- $E[S] = S_0 e^{\mu + \sigma^2/2}$
- $P(S > K) = P(Z > \ln(K/S_0)) = 1 - \Phi\left(\frac{\ln(K/S_0) - \mu}{\sigma}\right)$

For $E[e^{-r}S]$ with $Z \sim N(r - \sigma^2/2, \sigma^2)$:
$E[e^{-r}S] = e^{-r}S_0 E[e^Z] = e^{-r}S_0 e^{(r-\sigma^2/2)+\sigma^2/2} = S_0$

Template E: Bayesian Discrete Prior

When you see: "prior", "posterior", "defective rate"

Steps:

1. **List hypotheses:** $\theta_1, \theta_2, \ldots$
2. **Priors:** $P(\theta_i)$
3. **Likelihoods:** $P(\text{data}|\theta_i)$
4. **Bayes:** $P(\theta_i|\text{data}) = \frac{P(\text{data}|\theta_i)P(\theta_i)}{\sum_j P(\text{data}|\theta_j)P(\theta_j)}$
5. **Normalize:** Make sure posteriors sum to 1

Template F: Bivariate Normal Conditional

When you see: "bivariate normal", "$Y|X = x$", "conditional distribution", "given $X = x$"

Main Formula: $\boxed{Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1 - \rho^2)\sigma_Y^2\right)}$

Step-by-Step Procedure:

1. **Identify parameters:** $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$
2. **Conditional mean:** $E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$
3. **Conditional variance:** $\text{Var}(Y|X) = \sigma_Y^2(1 - \rho^2)$ (constant in x!)
4. **Write distribution:** $Y|X = x \sim N(\text{cond. mean}, \text{cond. var})$

Worked Example: (X, Y) bivariate normal with $\mu_X = 0$, $\mu_Y = 0$, $\sigma_X = 1$, $\sigma_Y = 2$, $\rho = 0.6$.

- Find $Y|X = 3$:
- Cond. mean: $0 + 0.6 \cdot \frac{2}{1} \cdot (3 - 0) = 3.6$
- Cond. variance: $4 \cdot (1 - 0.36) = 4 \cdot 0.64 = 2.56$
- $Y|X = 3 \sim N(3.6, 2.56)$
- $P(Y > 5|X = 3) = P(Z > \frac{5-3.6}{1.6}) = P(Z > 0.875) \approx 0.19$

Special Cases:

- $\rho = 0$: $Y|X = x \sim N(\mu_Y, \sigma_Y^2)$ (independent, conditioning doesn't help)
- $\rho = \pm 1$: $\text{Var}(Y|X) = 0$ (perfect prediction, Y determined by X)

Template G: Linear Combination Independence

When you see: "$Y_1 = aX_1 + X_2$", "$Y_2 = X_1 + bX_2$", "independent components"

Steps:

1. **Setup:** X_1, X_2 i.i.d. $N(0,1)$
2. **Key insight:** For linear combinations of Gaussians, independence $\Leftrightarrow$ zero covariance
3. **Calculate:** $\text{Cov}(Y_1, Y_2) = a \cdot 1 + b \cdot 1 = a + b$
4. **Solve:** $a + b = 0 \Rightarrow b = -a$
5. **Joint density:** Product of marginals (since independent)

Template H: Predictive Distributions (Bayesian)

When you see: "predict next outcome", "predictive probability", "posterior predictive"

Steps:

1. **Prior predictive:** $P(X_{n+1} = x) = \sum_\theta P(X = x|\theta)P(\theta)$
2. **Posterior predictive:** $P(X_{n+1} = x|\text{data}) = \sum_\theta P(X = x|\theta)P(\theta|\text{data})$
3. **Use:** Posterior from Bayesian update in Step 2

Example: Dice problem: Prior $P(\theta)$ for die type, observe data, predict next roll.

Template I: Product of Lognormals

When you see: "XY where X, Y are lognormal", "product of independent"

Key Insight: $\ln(XY) = \ln X + \ln Y$

Steps:

1. If $\ln X \sim N(\mu_1, \sigma_1^2)$ and $\ln Y \sim N(\mu_2, \sigma_2^2)$ independent
2. Then $\ln(XY) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
3. So XY is lognormal with parameters $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
4. $E[XY] = E[X]E[Y]$ (independence) $= e^{\mu_1+\sigma_1^2/2} \cdot e^{\mu_2+\sigma_2^2/2}$

Template J: Monty Hall Variants

When you see: "Monty Hall", "contestant picks", "host opens"

Steps:

1. **Define hypotheses:** H_A, H_B, H_C = car behind door A, B, C
2. **Priors:** Usually uniform $P(H_i) = 1/3$
3. **Key:** Likelihoods depend on host behavior!
   - **Sober:** Knows car location, opens non-car door
   - **Dizzy:** Opens random door (50-50)
4. **Sober Monty:** Switch doubles probability (2/3 vs 1/3)
5. **Dizzy Monty:** No advantage to switching

Template K: Finding n for CLT Probability

When you see: "smallest n such that", "how many samples needed"

Steps:

1. **Setup:** Want $P(\bar{X} > c) > p$ or $P(\bar{X} < c) > p$
2. **CLT:** $\bar{X} \approx N(\mu, \sigma^2/n)$
3. **Standardize:** $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > z\right) = \text{target}$
4. **Solve for** n**:** Using z-table, find z^* then solve $\frac{c-\mu}{\sigma/\sqrt{n}} = z^*$
5. **Result:** $n \geq \left(\frac{z^*\sigma}{c-\mu}\right)^2$

Template L: Max/Min of i.i.d. Variables

When you see: "maximum of n", "minimum of n", "largest/smallest"

Formulas:

- $P(\max < a) = P(\text{all} < a) = [F(a)]^n$ (if i.i.d.)
- $P(\max > a) = 1 - [F(a)]^n$
- $P(\min < a) = 1 - [1 - F(a)]^n$
- $P(\min > a) = [1 - F(a)]^n$

For Uniform(0,1): $E[X_{(n)}] = \frac{n}{n+1}$, $E[X_{(1)}] = \frac{1}{n+1}$

Template M: Conditioning on Event

When you see: "given that $X > a$", "conditional on event"

Steps:

1. **Conditional CDF:** $P(X \leq x|X > a) = \frac{P(a<X\leq x)}{P(X>a)}$ for $x > a$
2. **For Exponential:** Memoryless! $P(X > s + t|X > s) = P(X > t)$
3. **General:** Use $f_{X|A}(x) = f_X(x)/P(A)$ for $x \in A$

Template N: Bivariate Normal from Conditions

When you see: "$E[Y|X = x] = ...$", "$\text{Var}(Y|X) = ...$", "find parameters"

Key Formulas:

- $E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$
- $\text{Var}(Y|X) = \sigma_Y^2(1 - \rho^2)$ (constant!)

Method: Match coefficients to extract μ_Y, $\rho\frac{\sigma_Y}{\sigma_X}$, and $\sigma_Y^2(1 - \rho^2)$

Template O: Probability Involving Sample Average

When you see: "$P(\bar{X}/(\bar{X} + c) < p)$", "ratio with sample mean"

Steps:

1. **Transform:** Simplify inequality algebraically first!
2. **Example:** $\frac{\bar{X}}{\bar{X}+3} < 0.5 \Leftrightarrow \bar{X} < 3$
3. **Apply CLT:** $\bar{X} \approx N(\mu, \sigma^2/n)$
4. **Calculate:** Standard normal probability

Template P: Sum of Independent Poissons

When you see: "sum of Poisson", "combined arrivals"

Key Property: If $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$ independent: $\boxed{X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)}$

Method: Use MGF: $M_{X+Y}(t) = e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$

APPENDIX A: COMPLETE FORMULA SHEET

Probability Formulas

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A^c) = 1 - P(A)$
- $P(A|B) = P(A \cap B)/P(B)$
- $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$
- $P(A) = \sum P(A|B_i)P(B_i)$ (Total Probability)
- $P(H|E) = P(E|H)P(H)/P(E)$ (Bayes)
- $C(n,k) = n!/(k!(n-k)!)$
- $P(n,k) = n!/(n-k)!$
- Independent: $P(A \cap B) = P(A)P(B)$
- DeMorgan: $(A \cup B)^c = A^c \cap B^c$

Expectation & Variance

- $E[X] = \sum xp(x)$ or $\int xf(x)dx$
- $E[g(X)] = \sum g(x)p(x)$ or $\int g(x)f(x)dx$
- $E[aX+b] = aE[X]+b$
- $E[X+Y] = E[X] + E[Y]$ (always!)
- $E[XY] = E[X]E[Y]$ (only if independent)
- $\text{Var}(X) = E[X^2] - (E[X])^2$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$

Conditional Expectation

- $E[X|Y = y] = \sum xP(X = x|Y = y)$ or $\int xf(x|y)dx$
- $E[X] = E[E[X|Y]]$ (Total Expectation)
- $E[h(Y)X|Y] = h(Y)E[X|Y]$
- $\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$
- If indep: $E[X|Y] = E[X]$

MGF & Limit Theorems

- $M_X(t) = E[e^{tX}]$
- $E[X^k] = M^{(k)}(0)$
- $M_{X+Y}(t) = M_X(t)M_Y(t)$ (if independent)
- $M_{aX+b}(t) = e^{bt}M_X(at)$
- CLT: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$
- For sum: $S_n \approx N(n\mu, n\sigma^2)$
- For mean: $\bar{X} \approx N(\mu, \sigma^2/n)$
- CI: $\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$

Bivariate Normal

- $E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$
- $\text{Var}(Y|X) = \sigma_Y^2(1 - \rho^2)$
- $\rho = 0 \Leftrightarrow$ independent (MVN only!)
- $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 +$

APPENDIX B: DISTRIBUTION CHEAT SHEET

Distribution	PMF/PDF	Mean	Variance	MGF	Notes
Discrete Distributions					
Bernoulli(p)	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$(1-p)+pe^t$	$x \in \{0,1\}$
Binomial(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$	$(1-p+pe^t)^n$	$k = 0,...,n$
Poisson(λ)	$e^{-\lambda}\lambda^k/k!$	λ	λ	$e^{\lambda(e^t-1)}$	$k = 0,1,2,...$
Geometric(p)	$p(1-p)^k$	$(1-p)/p$	$(1-p)/p^2$	$p/(1-(1-p)e^t)$	Memoryless
Neg. Binomial(r,p)	$\binom{k+r-1}{k}p^r(1-p)^k$	$r(1-p)/p$	$r(1-p)/p^2$	$(p/(1-(1-p)e^t))^r$	–
Hypergeometric	Complex	nK/N	Complex	–	No replacement
Continuous Distributions					
Uniform(a,b)	$1/(b-a)$	$(a+b)/2$	$(b-a)^2/12$	$(e^{tb}-e^{ta})/(t(b-a))$	
Normal(μ,σ^2)	$(2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2	$e^{\mu t+\sigma^2t^2/2}$	–
Exponential(λ)	$\lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$	$\lambda/(\lambda-t)$	Memoryless
Gamma(r,λ)	$\lambda^r x^{r-1}e^{-\lambda x}/\Gamma(r)$	r/λ	r/λ^2	$(\lambda/(\lambda-t))^r$	–
Beta(α,β)	$x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta)$	$\alpha/(\alpha+\beta)$	Complex	–	$x \in (0,1)$

APPENDIX C: PROFESSOR'S NOTATION GUIDE

- Uses $\psi(t)$ for MGF (not $M(t)$)
- Writes $\text{Var}(X)$ not σ_X^2
- Uses $f(x)$ for both PMF and PDF
- $g_1(x|y)$ for conditional PDF of $X|Y$
- $\pi(\theta)$ for prior, $\pi(\theta|x)$ for posterior
- $L(x|\theta)$ for likelihood
- H_i for hypotheses in Bayes problems
- $\bar{X}$ or $\bar{X}_n$ for sample mean
- $X_{(k)}$ for k-th order statistic
- I_A for indicator of event A
- $\xrightarrow{d}$ for convergence in distribution
- $\xrightarrow{P}$ for convergence in probability
- $\Phi(z)$ for standard normal CDF
- z_α for quantile where $P(Z > z_\alpha) = \alpha$
- Finance: S_t for stock price at time t

APPENDIX D: LAST-MINUTE REVIEW CHECKLIST

Time Management (90 minutes, 3 questions)

- ☐ **0-5 min:** Read all problems, identify types using decision tree
- ☐ **5-35 min:** Question 1 (aim for 30 min max)
- ☐ **35-65 min:** Question 2 (aim for 30 min max)
- ☐ **65-85 min:** Question 3 (aim for 20 min)
- ☐ **85-90 min:** Review, check arithmetic

High-Yield Topics to Review (Post-M2 Focus)

1. *** Central Limit Theorem applications**
2. *** Bivariate Normal problems**
3. *** Bayesian updates (especially Monty Hall variants)**
4. *** Conditional Expectation and Total Expectation**
5. *** Lognormal/Finance applications**
6. Joint distributions (finding marginals, checking independence)
7. Covariance and correlation calculations
8. MGF for finding distributions of sums
9. Normal approximations with continuity correction
10. Confidence intervals using CLT

What to Memorize vs Look Up

MEMORIZE:

- "Gaussian" = Normal, "Gaussian vector" = MVN
- Normal standardization: $Z = (X - \mu)/\sigma$

- CLT: $(\bar{X}-\mu)/(\sigma/\sqrt{n}) \to N(0,1)$
- Lognormal: $E[e^X] = e^{\mu+\sigma^2/2}$ for $X \sim N(\mu,\sigma^2)$
- BVN Conditional: $\mu_{Y|X} = \mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X)$
- For MVN ONLY: $\rho = 0 \Leftrightarrow$ independent

LOOK UP:

- Distribution tables (PMF/PDF formulas)
- MGF formulas: $\psi(t)$ values
- Jacobian details
- Normal table (Φ values)

PARAMETER TRAP CHECKLIST

- ☐ "Mean $\theta = 3$" (Exp) $\Rightarrow \lambda = 1/3$
- ☐ "Rate $\lambda = 2$" $\Rightarrow$ Mean $= 1/2$
- ☐ Check: Is it Geom(failures) or Geom(trials)?
- ☐ BVN: Is variance σ^2 or std dev σ?

Common Professor Patterns

- Part (a): Basic setup/calculation
- Part (b): Extension requiring part (a)
- Part (c): Conceptual twist or limiting behavior
- Finance context in at least one problem
- One Bayesian problem guaranteed
- One CLT/approximation problem guaranteed

Final Tips

- **! Always check if variables are independent before using simplified formulas**
- **! Apply continuity correction for discrete $\to$ continuous**
- **! Verify bounds of integration match the region**
- Start with problems you recognize immediately
- Show all work - partial credit is generous
- If stuck, write down relevant formulas and what you know
- Check units/reasonableness of final answers

Last-Minute Reminders

- **Exp mean trap:** Mean $\theta = 3 \Rightarrow \lambda = 1/3$
- **Normal notation:** $N(\mu,\sigma^2)$ uses variance, not std dev
- **Standardize:** $Z = (X-\mu)/\sigma$, NOT $(X-\mu)/\sigma^2$
- **CLT:** $\text{Var}(\bar{X}) = \sigma^2/n$, $\text{Var}(S_n) = n\sigma^2$
- **MVN only:** $\rho = 0 \Leftrightarrow$ independent
- **Jacobian:** Use ABSOLUTE VALUE $|J|$
- **BVN Conditional:** Variance doesn't depend on x!
- **Bayesian:** Posterior $\propto$ Likelihood $\times$ Prior
- **Lognormal:** $E[e^X] = e^{\mu+\sigma^2/2}$
- **Max CDF:** $F_{\max}(x) = [F(x)]^n$
- **Min survival:** $P(\min > x) = [1-F(x)]^n$

Problem Approach Strategy

1. Read problem carefully - identify KEY WORDS
2. Match to template or section using decision tree
3. Write down given information and what's asked
4. Write relevant formulas BEFORE computing
5. Show all steps (partial credit!)
6. Check: Is answer reasonable? Between 0 and 1 for probabilities?

APPENDIX E: PROPERTIES OF E, Var, Cov, SD

Expected Value Properties

- **Linearity (ALWAYS holds):**
  - $E[aX+b] = aE[X]+b$
  - $E[X+Y] = E[X]+E[Y]$ (no independence needed!)
  - $E[aX+bY+c] = aE[X]+bE[Y]+c$
  - $E[\sum_{i=1}^n X_i] = \sum_{i=1}^n E[X_i]$
- **Constants:**
  - $E[c] = c$ (constant has expected value = itself)
  - $E[cX] = cE[X]$
- **Product (ONLY if independent):**
  - $E[XY] = E[X]E[Y]$ **only if** X, Y **independent!**
  - $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ if independent
  - **! If dependent:** $E[XY] \neq E[X]E[Y]$ **in general!**
- **LOTUS (Law of the Unconscious Statistician):**
  - $E[g(X)] = \sum_x g(x)P(X=x)$ (discrete)
  - $E[g(X)] = \int g(x)f(x)dx$ (continuous)
  - Don't need distribution of $g(X)$, just distribution of X!

Variance Properties

! SCALARS SQUARE when factored out of Var!

- **Computational Formula:** $\boxed{\text{Var}(X) = E[X^2] - (E[X])^2}$
- **Linear Transform:** $\boxed{\text{Var}(aX+b) = a^2\text{Var}(X)}$
  - **! NOT $a\text{Var}(X)$! The constant SQUARES!**
  - $\text{Var}(2X) = 4\text{Var}(X)$, NOT $2\text{Var}(X)$
  - $\text{Var}(X+5) = \text{Var}(X)$ (adding constant doesn't change variance)
  - $\text{Var}(-X) = \text{Var}(X)$
- **Constants:**
  - $\text{Var}(c) = 0$ (constant has no variance)
  - $\text{Var}(X+c) = \text{Var}(X)$

- **Sum of Random Variables:** $\boxed{\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)}$ $\boxed{\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y)}$
  - If independent: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
  - If independent: $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)$ (still +!)
- **General Linear Combination:** $\boxed{\text{Var}(aX+bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)}$
  - If independent: $\text{Var}(aX+bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$
- **Sum of i.i.d.:** $\boxed{\text{Var}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2}$
- **Sample Mean:** $\boxed{\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}}$
- **Non-negativity:**
  - $\text{Var}(X) \geq 0$ always
  - $\text{Var}(X) = 0$ iff X is constant

Standard Deviation Properties

- **Definition:** $\sigma_X = \text{SD}(X) = \sqrt{\text{Var}(X)}$
- **Linear Transform:** $\boxed{\text{SD}(aX+b) = |a| \cdot \text{SD}(X)}$
  - $\text{SD}(2X) = 2 \cdot \text{SD}(X)$ (absolute value of scalar)
  - $\text{SD}(-X) = \text{SD}(X)$
  - $\text{SD}(X+5) = \text{SD}(X)$
- **Sample Mean:** $\boxed{\text{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}}$
  - Called "Standard Error" of the mean
- **! SD does NOT have nice additivity: SD$(X+Y) \neq$ SD(X) + SD(Y)**
- **For independent:**
  - $\text{SD}(X+Y) = \sqrt{\text{Var}(X) + \text{Var}(Y)} = \sqrt{\sigma_X^2 + \sigma_Y^2}$
  - Must go through variance, then take square root!

Covariance Properties

- **Computational Formula:** $\boxed{\text{Cov}(X,Y) = E[XY] - E[X]E[Y]}$
- **Definition Form:** $\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$
- **Symmetry:** $\text{Cov}(X,Y) = \text{Cov}(Y,X)$
- **Self-Covariance:** $\text{Cov}(X,X) = \text{Var}(X)$
- **Linear Transform:** $\boxed{\text{Cov}(aX+b, cY+d) = ac \cdot \text{Cov}(X,Y)}$
  - Adding constants doesn't affect covariance
  - Scalars multiply (not square!)
- **Linearity in Each Argument:**
  - $\text{Cov}(X+Y, Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$
  - $\text{Cov}(aX, Y) = a\text{Cov}(X,Y)$
- **Independence:**
  - If independent: $\text{Cov}(X,Y) = 0$
  - **! Converse FALSE: Cov$(X,Y) = 0$ does NOT imply independence!**
  - **Exception:** For MVN, Cov $= 0 \Leftrightarrow$ independent
- **With Constants:**
  - $\text{Cov}(X,c) = 0$ for any constant c
  - $\text{Cov}(X+c, Y) = \text{Cov}(X,Y)$

Correlation Properties

- **Definition:** $\boxed{\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}}$
- **Bounds:** $-1 \leq \rho \leq 1$
- **Linear Transform (Scale Invariant):** $\boxed{\rho_{aX+b,cY+d} = \text{sign}(ac) \cdot \rho_{XY}}$
  - If $a, c > 0$: $\rho_{aX+b,cY+d} = \rho_{XY}$
  - If $a > 0, c < 0$: $\rho_{aX+b,cY+d} = -\rho_{XY}$
- **Interpretation:**
  - $\rho = 1$: Perfect positive linear relationship
  - $\rho = -1$: Perfect negative linear relationship
  - $\rho = 0$: No linear relationship (uncorrelated)
- **Self-Correlation:** $\rho_{XX} = 1$

Key Relationships Summary

Operation	**E[]**	**Var()**
$aX+b$	$aE[X]+b$	$a^2\text{Var}(X)$
$X+Y$	$E[X]+E[Y]$	$\text{Var}(X)+\text{Var}(Y)+2\text{Cov}$
$X+Y$ (indep)	$E[X]+E[Y]$	$\text{Var}(X)+\text{Var}(Y)$
$\sum X_i$ (i.i.d.)	$n\mu$	$n\sigma^2$
$\bar{X}$ (i.i.d.)	μ	σ^2/n

Common Mistakes to Avoid

☐ $\text{Var}(2X) = 2\text{Var}(X)$ **WRONG!** Correct: $4\text{Var}(X)$

☐ $\text{SD}(X+Y) = \text{SD}(X) + \text{SD}(Y)$ **WRONG!** (no additivity for SD)

☐ $E[XY] = E[X]E[Y]$ always **WRONG!** (only if independent)

☐ $\text{Var}(X-Y) = \text{Var}(X) - \text{Var}(Y)$ **WRONG!** It's + not -

☐ $\text{Cov}(X,Y) = 0 \Rightarrow$ independent **WRONG!** (only for MVN)

- ☐ $\text{Var}(\bar{X}) = \sigma^2$ **WRONG!** Correct: σ^2/n
- ☐ $\text{Var}(S_n) = \sigma^2/n$ **WRONG!** Correct: $n\sigma^2$

Covariance Matrix Definition

- **For 2D vector** (X, Y): $\boxed{\boldsymbol{\Sigma} = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X,Y) \\ \text{Cov}(X,Y) & \text{Var}(Y) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}}$
- **For n-dim vector $\mathbf{X} = (X_1, ..., X_n)$:** $\boxed{\boldsymbol{\Sigma}_{ij} = \text{Cov}(X_i, X_j)}$
  - Diagonal: $\Sigma_{ii} = \text{Var}(X_i)$
  - Off-diagonal: $\Sigma_{ij} = \text{Cov}(X_i, X_j)$
  - Symmetric: $\Sigma_{ij} = \Sigma_{ji}$
- **Independent components:** $\boldsymbol{\Sigma}$ is diagonal (off-diagonal = 0)

How to Find $E[X^2]$

- **From Variance Formula:** Rearranging $\text{Var}(X) = E[X^2] - (E[X])^2$: $\boxed{E[X^2] = \text{Var}(X) + (E[X])^2}$
- **Step-by-Step:**
  1. Find $E[X]$ (mean)
  2. Find $\text{Var}(X)$ (from distribution or given)
  3. Calculate: $E[X^2] = \text{Var}(X) + (E[X])^2$
- **Example:** $X \sim N(3, 4)$ (mean 3, variance 4)
  - $E[X] = 3$, $\text{Var}(X) = 4$
  - $E[X^2] = 4 + 3^2 = 4 + 9 = 13$
- **Example:** $X \sim \text{Exp}(\lambda = 2)$
  - $E[X] = 1/2$, $\text{Var}(X) = 1/4$
  - $E[X^2] = 1/4 + (1/2)^2 = 1/4 + 1/4 = 1/2$

How to Calculate Standard Deviation

- **From Variance:** $\boxed{\text{SD}(X) = \sigma = \sqrt{\text{Var}(X)}}$
- **Step-by-Step from Data/Distribution:**
  1. Find $E[X]$ (mean)
  2. Find $E[X^2]$
  3. Compute $\text{Var}(X) = E[X^2] - (E[X])^2$
  4. Take square root: $\text{SD}(X) = \sqrt{\text{Var}(X)}$
- **For named distributions:** Use formula table
  - $\text{Normal}(\mu, \sigma^2)$: $\text{SD} = \sigma$
  - $\text{Exp}(\lambda)$: $\text{SD} = 1/\lambda$
  - $\text{Binomial}(n, p)$: $\text{SD} = \sqrt{np(1-p)}$
  - $\text{Poisson}(\lambda)$: $\text{SD} = \sqrt{\lambda}$

Joint Density of a Gaussian Vector

- **Bivariate Normal** (X, Y): $\boxed{f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{Q}{2(1-\rho^2)}\right)}$

  where $Q = \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}$
- **When $\rho = 0$ (independent):** $f(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2} -\right.$
- **Matrix Form (General MVN):** $f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}\right.$
- **Key Insight:** For independent components, joint = product of marginals

How to Find Conditional Distribution

General Method (Any Joint Distribution):

1. Find joint PDF: $f(x, y)$
2. Find marginal of what you're conditioning on: $f_Y(y) = \int f(x,y)\,dx$
3. Apply formula: $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$
4. Simplify and identify distribution

For Bivariate Normal (Use Formulas Directly!):

- **Conditional Distribution:** $\boxed{X|Y = y \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1-\rho^2)\right)}$
- **Step-by-Step:**
  1. Identify: $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$
  2. Conditional mean: $E[X|Y = y] = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$
  3. Conditional variance: $\text{Var}(X|Y) = \sigma_X^2(1-\rho^2)$
  4. Write: $X|Y = y \sim N(\text{cond mean}, \text{cond var})$
- **Worked Example:** (X, Y) bivariate normal: $\mu_X = 2$, $\mu_Y = 5$, $\sigma_X = 3$, $\sigma_Y = 4$, $\rho = 0.5$
  - Find $X|Y = 9$:
  - Cond. mean: $2 + 0.5 \cdot \frac{3}{4}(9 - 5) = 2 + 0.375 \cdot 4 = 3.5$
  - Cond. var: $9(1 - 0.25) = 9 \cdot 0.75 = 6.75$
  - $X|Y = 9 \sim N(3.5, 6.75)$

For Discrete (Table Method):

1. Fix the conditioning value (e.g., $Y = y$)
2. Take the row/column for that value from joint table
3. Divide each entry by the marginal $P(Y = y)$
4. Result is conditional PMF $P(X = x|Y = y)$