STAT 131 — Midterm Review (Practice Set)

Prepared by Antonio Aguirre

Important note. This practice set targets the main themes on the midterm, but it is *not* a complete checklist. You should still review your lecture notes, homework, and discussion work to make sure you have full coverage. Also, this is not a comprehesnive summary for all the formulas you will need.

Quick recap

Conditioning, complements, and Bayes

- **Conditional probability (definition):** $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$ for $P(B) > 0$.
- **Product rule (same statement, different view):** $P(A \cap B) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B)$.
- **Complements:** $P(A^c) = 1 - P(A)$.
- **"At least one" trick:** $P(\text{at least one}) = 1 - P(\text{none})$.
- **Law of total probability (partition $\{B_1, \ldots, B_m\}$):** $P(A) = \sum_{i=1}^{m} P(A \mid B_i)\,P(B_i)$.
- **Bayes' rule:** $P(B_j \mid A) = \dfrac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{m} P(A \mid B_i)P(B_i)}$.

Discrete models you are implicitly using here

- **Binomial setting:** X counts "successes" in n independent trials with success probability p.
- **Hypergeometric setting:** sampling *without replacement*; probabilities are combinations:

$$P(C = k) = \frac{\binom{(\#\text{ success items})}{k}\binom{(\#\text{ non-success items})}{n-k}}{\binom{(\#\text{ total items})}{n}}.$$

Joint distributions: discrete vs. continuous

- **Discrete joint pmf (table):** $p_{X,Y}(x, y) = P(X = x, Y = y)$ and $\sum_x \sum_y p_{X,Y}(x, y) = 1$.
- **Continuous joint/one-dimensional pdf:** densities are *not* probabilities. In particular $P(X = x) = 0$ for any single point.
- **Get probabilities:** discrete $\Rightarrow$ *sum* over points; continuous $\Rightarrow$ *integrate* over a region/interval.
- **Marginals (sum/integrate out):** $p_X(x) = \sum_y p_{X,Y}(x, y),\ \ f_X(x) = \int f_{X,Y}(x, y)\,dy$.
- **Conditionals (divide by the marginal):** $p_{X|Y}(x \mid y) = \dfrac{p_{X,Y}(x, y)}{p_Y(y)},\ \ f_{X|Y}(x \mid y) = \dfrac{f_{X,Y}(x, y)}{f_Y(y)}$.
- **Independence test (both cases):** $X \perp Y \iff$ joint = product of marginals on the support.

CDFs and quantiles

- **CDF definition:** $F(x) = P(X \le x)$.
- **Interval probability from a CDF:** $P(a < X \le b) = F(b) - F(a)$ (works for any distribution).
- **p-quantile:** $q_p = \min\{x : F(x) \ge p\}$.

Problem 1

Over 6 days, Antonio tries to be responsible and carry a charger. The probability his laptop battery dies anyway is $p = 0.15$ each day, independently day-to-day. Let X be the number of days the battery dies.

(a) Compute $P(X \geq 1)$. *Hint: use a complement.*

(b) Compute $P(X \geq 2 \mid X \geq 1)$. *Hint: use* $P(A \mid B) = P(A \cap B)/P(B)$ *with* $A = \{X \geq 2\}$, $B = \{X \geq 1\}$.

Problem 2

In the city where Antonio lives, a "smart safety" analytics program rolls out by the local gov. Two vendors are piloting it:

- Vendor A (generic dashboard): handles 70% of neighborhoods. It triggers a "high-risk" flag on 4% of residents.
- Vendor B (Palantir): handles 30% of neighborhoods. It triggers a "high-risk" flag on 89% of residents.

Let $H =$ "Antonio gets labeled high-risk."

(a) Compute $P(H)$ using the law of total probability.

(b) Compute $P(\text{Vendor B} \mid H)$.

(c) (Sanity check) Should $P(\text{Vendor B} \mid H)$ be larger or smaller than 0.30? Explain in one sentence.

Hint: partition by which vendor handled the neighborhood.

Problem 3

Antonio attends a conference and collects some free candy. A bag has 9 chocolate candies and 6 fruit candies. Antonio draws 5 *without replacement*. Let C be the number of chocolate candies drawn.

(a) Compute $P(C = 3)$.

(b) Compute $P(C = 3 \mid C \geq 1)$.

Hint: for (b), use $P(C = 3 \mid C \geq 1) = \dfrac{P(C = 3)}{P(C \geq 1)}$ *and* $P(C \geq 1) = 1 - P(C = 0)$.

Problem 4

Antonio is doomscrolling a feed where the ratio of AI slop to real content is distressingly nonzero. Let $X \in \{0, 1, 2\}$ be the number of AI-generated images he sees in the next 3 posts, and let $Y \in \{0, 1\}$ indicate whether he gets a crypto ad in that batch ($Y = 1$ means yes). The joint pmf is:

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	0.10	0.18	0.12
$Y = 1$	0.20	0.25	0.15

(a) Compute $p_X(0), p_X(1), p_X(2)$ and $p_Y(0), p_Y(1)$.

(b) Compute $P(X \geq 1 \mid Y = 0)$ and $P(Y = 1 \mid X = 2)$.

(c) Are X and Y independent? Justify using one clear check of the form $p_{X,Y}(x, y)$ vs. $p_X(x)p_Y(y)$.

Hint: do marginals first; conditionals are "divide by the right marginal."

Problem 5

Antonio submits a customer-support ticket and measures the response time T (in days). A continuous random variable T has density

$$f_T(t) = \begin{cases} c\,t, & 0 < t < 1, \\ c\,(2-t), & 1 \le t < 2, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find c.

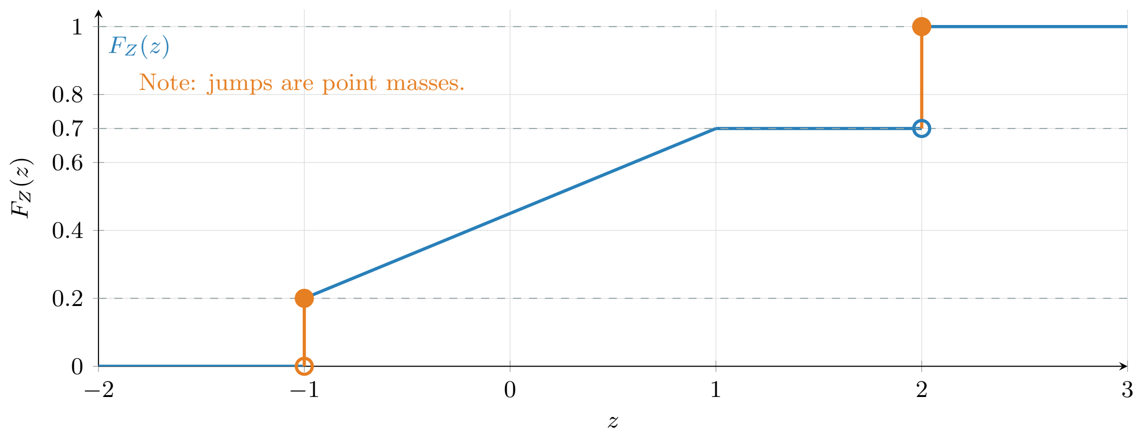
(b) Find the CDF $F_T(t)$ as a piecewise function.

(c) Compute $P(0.5 < T < 1.5)$.

Checkpoint: your CDF should be continuous and satisfy $F_T(2) = 1$.

Problem 6

Antonio is assigned an automated "trust score" Z by a black-box system to identify AI content. Most of the time, the score moves continuously, but occasionally the system makes a dramatic, unreviewable decision and snaps to specific values (those are the jumps). The plot below is the CDF $F_Z(z) = P(Z \leq z)$ of this mixed distribution (continuous part *and* point masses).



Answer the following (show brief reasoning; you can read values from the plot):

(a) Find $P(Z = -1)$ and $P(Z = 2)$. *Hint: point masses are jump sizes.*

(b) Compute $P(-1 < Z \leq 0.5)$. *Hint: use* $P(a < Z \leq b) = F(b) - F(a)$.

(c) Compute $P(Z > 1)$.

(d) Find the median (the 0.5 quantile) and the 0.9 quantile of Z using the definition:

$$q_p = \min\{z : \ F_Z(z) \geq p\}.$$