Stat C131A: Statistical Methods for Data Science

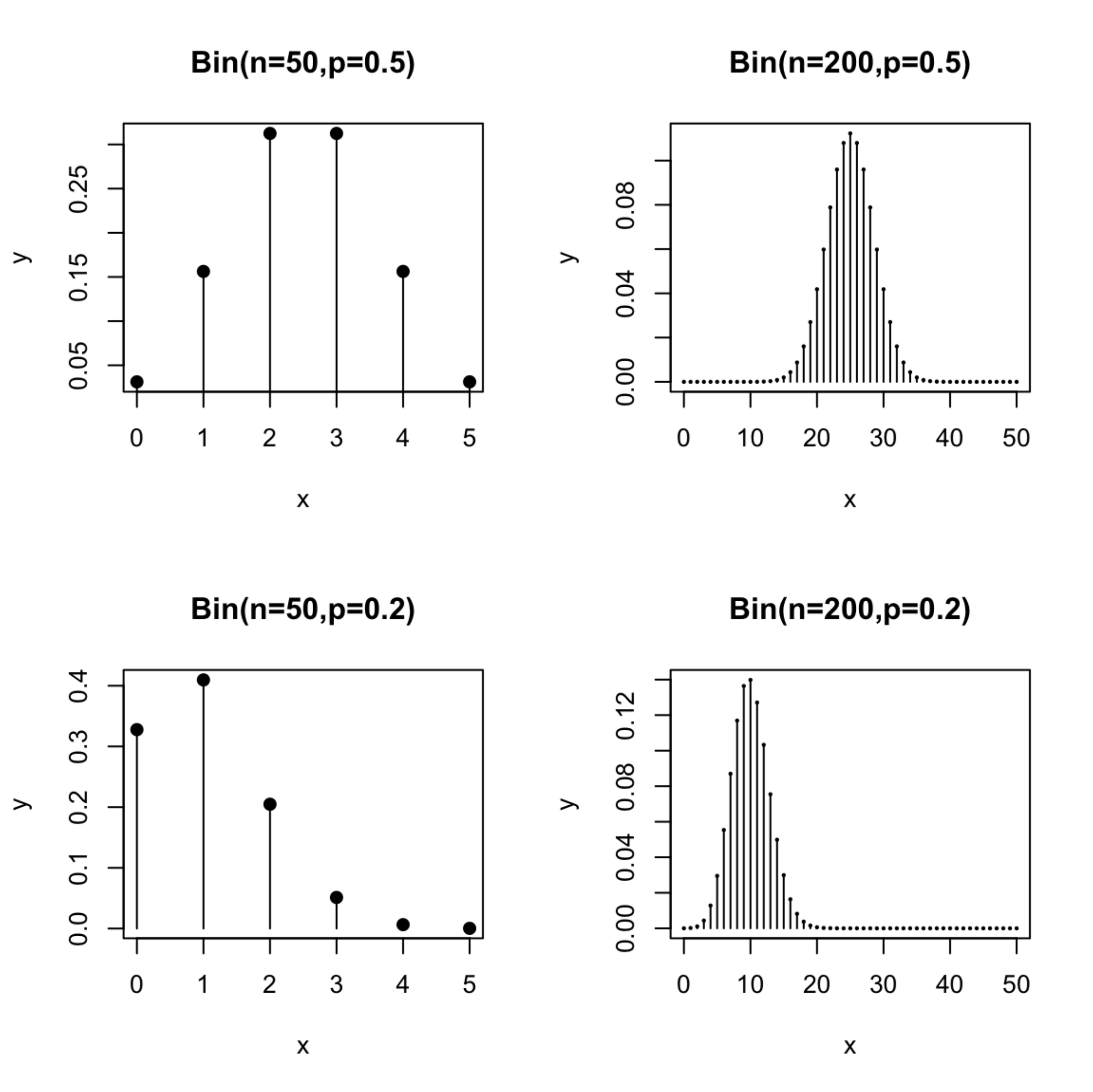
Lecture 5: Probability Distributions …contd

Sep 11 2025

Exams and Schedule

- We **will** have a written Final Exam
  - There will likely be a final project as well but **not in lieu** of a written final exam
- **Dates**
  - Midterm Exam (one only): **Thursday October 30th**
  - Final Exam: **Thursday December 18th**
  - Final project* submission: **Friday Dec 19th**

Coin Flips: Binomial distribution



- Experiment: n independent trials; each trial is success/failure with probability p of success.
- Parameter n: number of trials (not histogram bins).
- Parameter p: success probability per trial.
- x-axis: k = 0,1,2,...,n (every integer count is a point).
- y-axis: P(X = k)
- Note: Seeing **only a few** x-tick labels (such as 0-5 in top-left) is an axis choice; otherwise the pmf has n+1 discrete points.

Conditional Probability

- **Conditional** distribution
- Note one but **two** random events
- Independent events
  - $P(A \cap B) = P(A)P(B)$
- $P(A \cap B) = P(B)\, P(A \mid B)$
- $P(A \mid B) = P(A \cap B) \,/\, P(B)$
- **Bayes theorem**
  - $P(A \mid B) = P(B \mid A)\, P(A) \,/\, P(B)$
  - Where
    - $P(A)$ is the **prior**
    - $P(B \mid A)$ is the **likelihood**
    - $P(B)$ is to **normalize**
- **Independence** can be **Unconditional,** as well as **Conditional**

Probability Distributions: Expectation, Variance & Continuous Distributions

- **References**
  - **2.2.4*, 2.3.1, 2.3.3, 2.3.3***

Expectation

The **expectation** (or **mean**) of a distribution is defined as

$$E(X) = \sum_{k \in \Omega} k\, p(k)$$

$$E(X) = \sum_{k \in \Omega} k\, p(k) = 1P(X=1) + 2P(X=2) + \ldots + 6P(X=6)$$

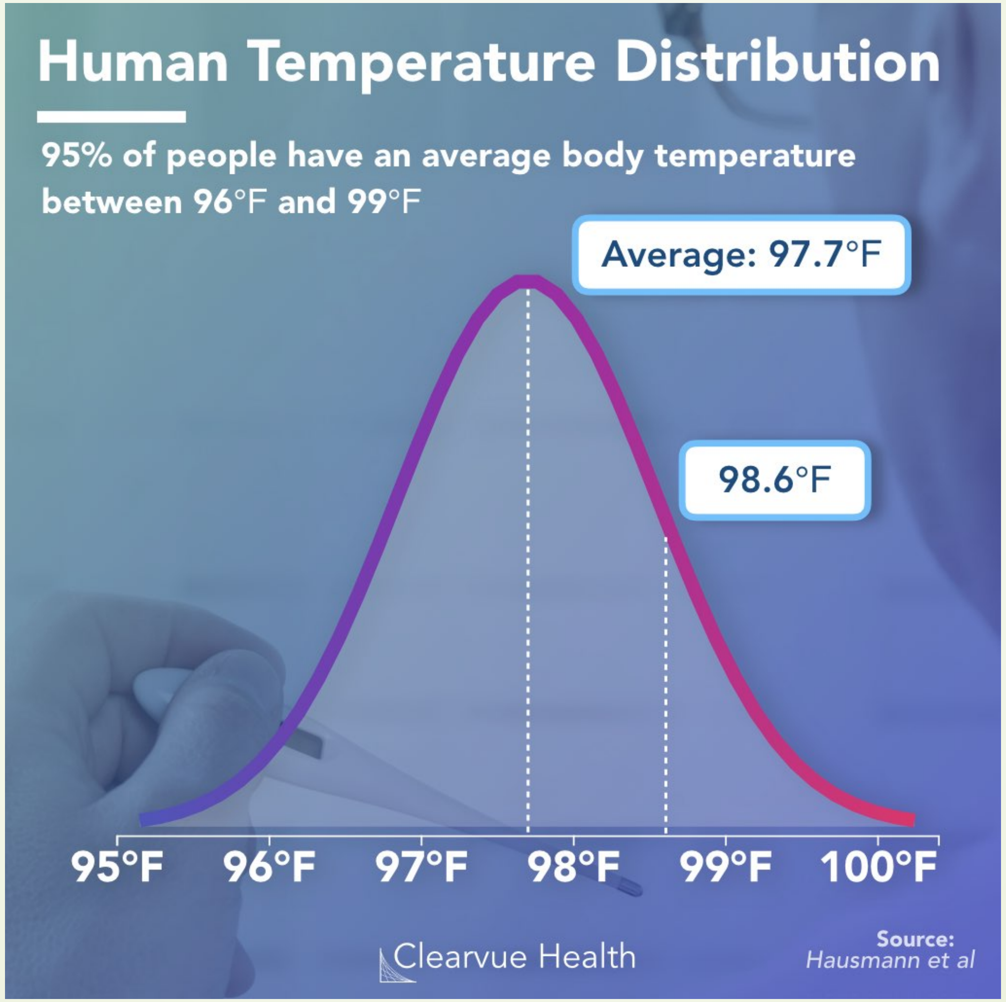
$$= 1/6(1 + 2 + \ldots + 6)$$

$$= 21/6 = 3.5$$

Expectation: Interpretation

- **Weighted mean** of outcomes
- Probability **acts as weight**

Expected Body Temperature ?



Variance

$$var(X) = E(X - E(X))^2 = \sum_{k \in \Omega} (k - E(X))^2 p(k)$$

$$\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 \frac{1}{n-1}$$

Expectation & Variance

- Expectation: **Center**
- Variance**: Spread**

Properties

- $E(a + bX) = a + bE(X)$
- $var(a + bX) = b^2 var(X)$
- $E(g(X)) \neq g(E(X))$; generally …
- $var(X) = E(X - E(X))^2 = E(X^2) - [E(X)]^2$

Discrete Distributions —> Continuous Distributions

- **Discrete:** Ω countable (dice, coin flips)
- **Continuous:** Ω interval (e.g. salaries $[0,\infty)$)

Key Idea: Infinite Outcomes

- Probability is **undefined for a single point** !
  - Cannot assign positive probability to each point

- **WHY ?**

Properties of Continuous Distributions

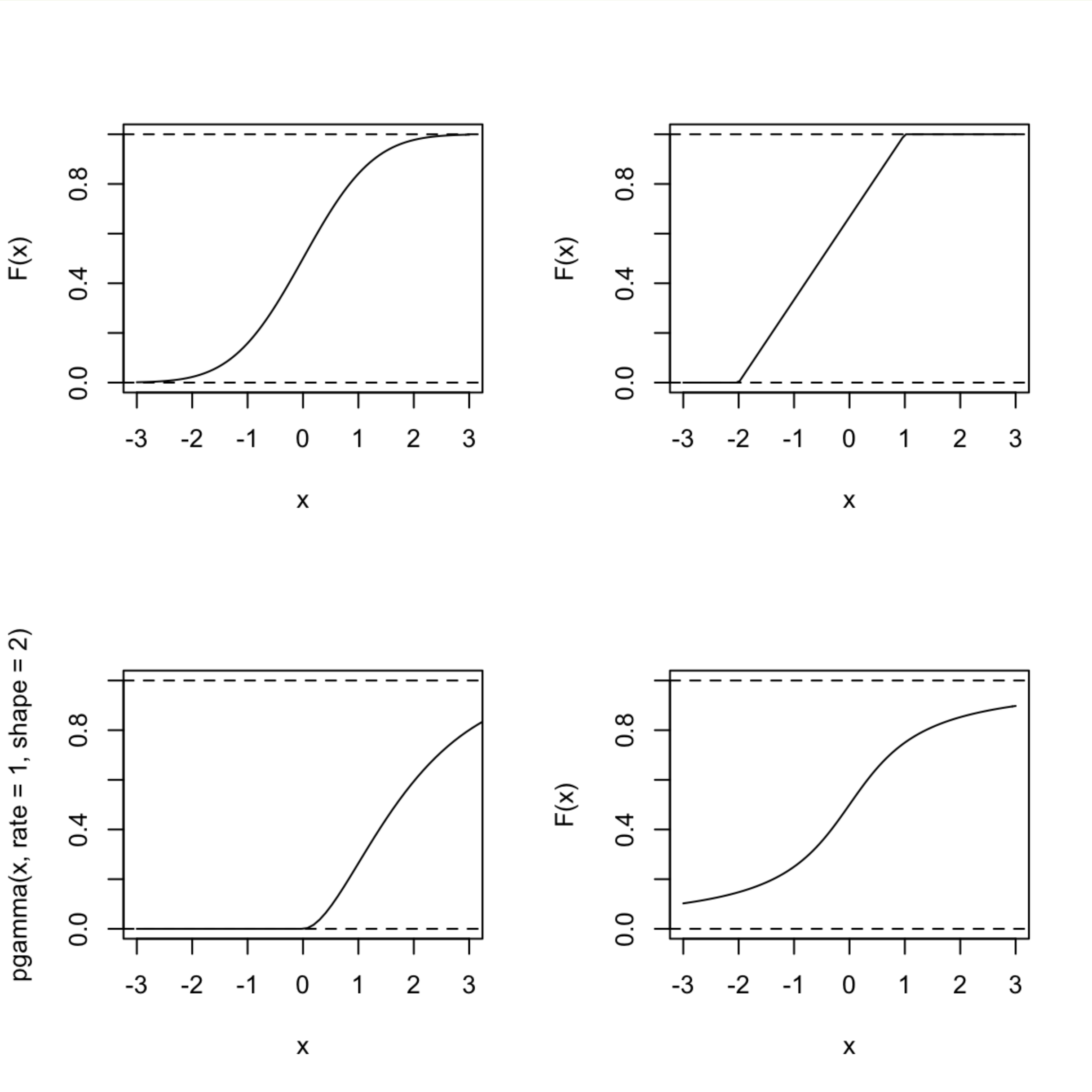
- $0 \leq P(A) \leq 1$, inclusive
- Probabilities are only calculated for events that are intervals, not individual points/outcomes
- $P(\Omega) = 1$

Probabilities over Intervals

- $P(a \leq X \leq b)$ can be > 0

Cumulative Distribution Function

- **Probability**
  - **Discrete** distributions: comes from **PMF**
  - **Continuous** distributions: comes from from **CDF** (a bit differently)
- $P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1)$
- $F(x) = P(X \leq x)$
- The function F is the cdf !


F(x)
0.0
0.4
0.8
-3
-2
-1
0
1
2
3
x
F(x)
0.0
0.4
0.8
-3
-2
-1
0
1
2
3
x
pgamma(x, rate = 1, shape = 2)
0.0
0.4
0.8
-3
-2
-1
0
1
2
3
x
F(x)
0.0
0.4
0.8
-3
-2
-1
0
1
2
3
x

Properties of CDFs

- **Non-decreasing**
- Limit as $x \rightarrow \infty$ = ?
- Limit as $x \rightarrow -\infty$ = ?

Properties of CDFs

- **Non-decreasing**
- Limit as $x \to \infty = 1$
- Limit as $x \to -\infty = ?$

Properties of CDFs

- **Non-decreasing**
- Limit as $x \to \infty = 1$
- Limit as $x \to -\infty = 0$

Why CDFs?

- The PMF does not apply in the continuous case
- The CDF **fully describes** the distribution

Interpreting CDFs

- $F(x)=0.8$ means ?

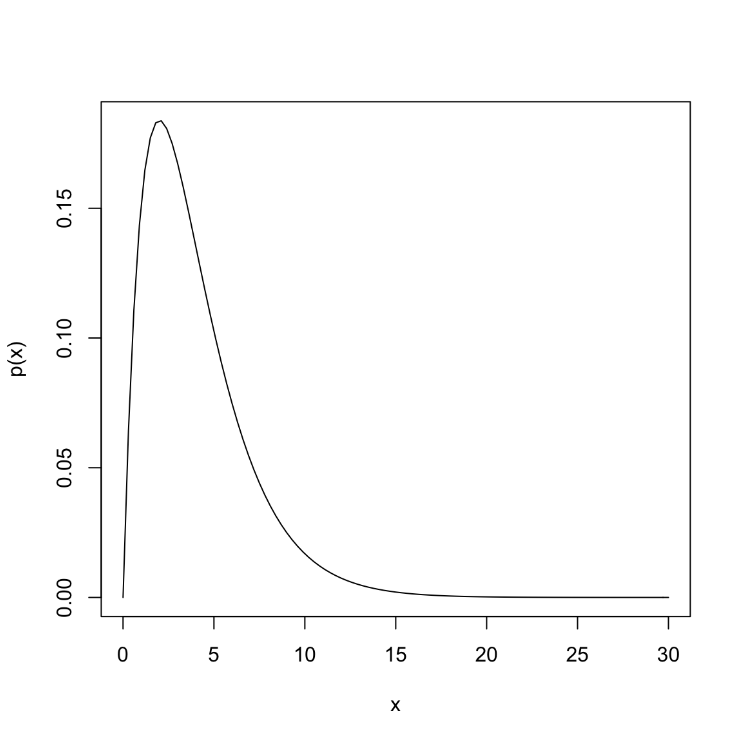
Interpreting CDFs

- $F(x)=0.8$ means 80% probability $\leq x$

The **Probability Density Function**

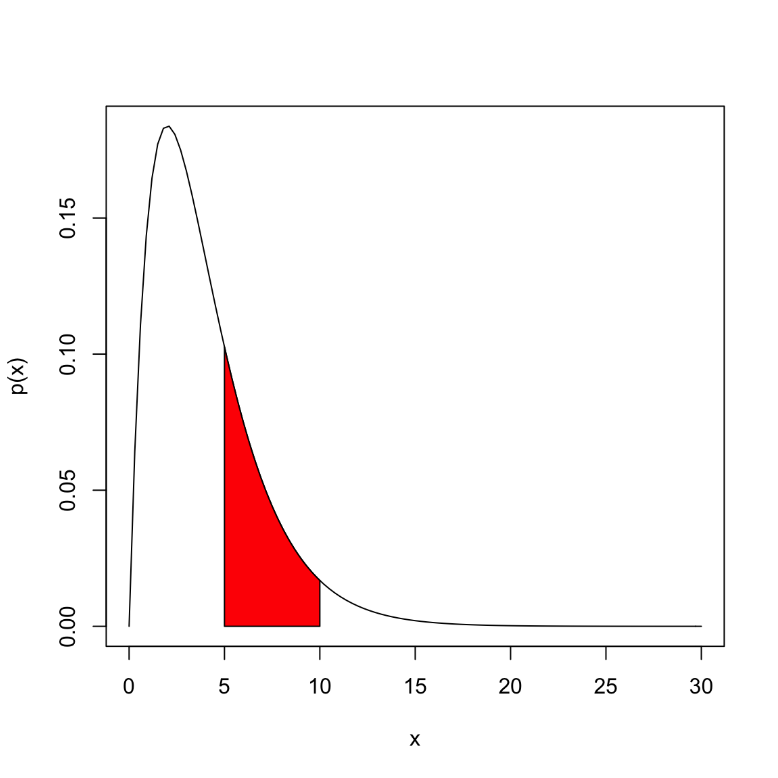
- Probability Density Function (PDF)
  - Like a Histogram
  - The analog of the PMF, for continuous distributions
- Formally, p(x) is the derivative of F(x); if F(X) is differentiable
  - $p(x) = \frac{d}{dx} F(x) \, .$
- Conversely, p(x) is that function such that if you take the area under its curve for the interval (a,b) it give us ?
  - The probability of that interval !
  - $\int_a^b p(x) = P(a \leq X \leq b) = F(b) - F(a)$
- Relationship
  - PDF = derivative of CDF
  - CDF = integral of PDF

$$p(x) = \frac{1}{4} x e^{-x/2}$$



$P(5 \leq X \leq 10)$

$$\int_5^{10} \frac{1}{4} x e^{-x/2}$$



PDF

- Practical use ?
- Computers can compute the area !

PDF Properties

- A probability density function gives the probability of any interval by taking the area under the curve
- The total area under the curve $p(x)$ must be exactly equal to 1
- Unlike probabilities, the value of $p(x)$ can be ≥ 1 !!!

Key Takeaways

- Expectation & variance summarize distributions
- Continuous distributions: probabilities over intervals
- CDF describes continuous distributions