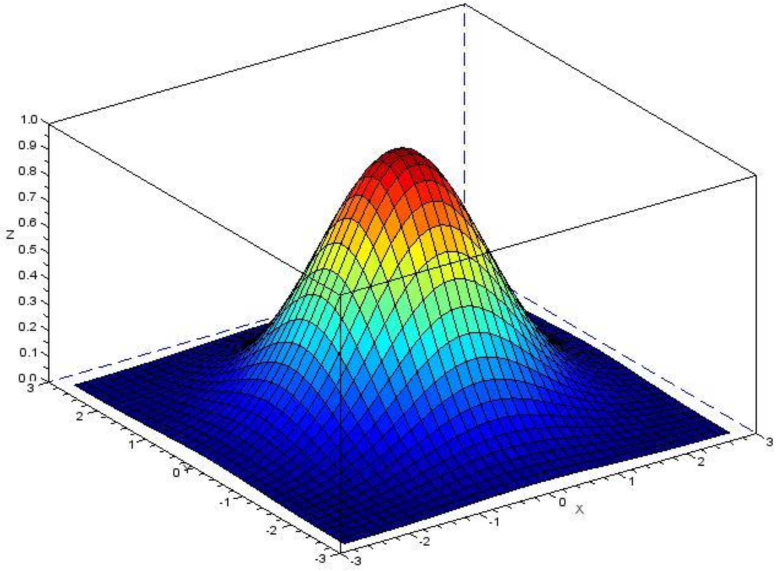
# CSE 312 Section 6



Slides made by Harrison Bay

# Administrivia

# Announcements & Reminders

- Midterm
  - Grades released on Gradescope – check your submission to read comments
  - Regrade requests open and close after a week
- HW4
  - Was due Tuesday 7/25 @ 11pm
  - Late deadline Thursday 7/27 @ 11pm
- HW5
  - Due Tuesday 8/1 @ 11pm

# Review & Questions

# Review of Main Concepts

- **Cumulative Distribution Function (CDF)**: For any random variable $X$, the CDF is defined as $F_X(x) = \mathbb{P}(X \leq x)$.
    - Notice the CDF is monotonic (non-decreasing), i.e. if $x < y$ then $F_X(x) \leq F_X(y)$
    - The CDF is bounded between 0 and 1.
- A **Continuous RV** is one for which the CDF is continuous everywhere.
    - Support is an uncountably infinite set.
- **Probability Density Function (PDF)** is defined as $f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$
    - Implies that $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t$
    - $\mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f_X(t)\mathrm{d}t$
    - $f_X(x) \geq 0$

# Review of Main Concepts

- **I.I.D**: Random variables are “independent and identically distributed” if they are independent and have the same PDF/PMF
- For continuous random variables:
  - $\mathbb{P}(X = x) = 0 \neq f_X(x)$
  - $F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t$
  - $\int_{-\infty}^{\infty} f_X(t)\mathrm{d}t = 1$
  - $\mathbb{E}[X] = \int_{-\infty}^{\infty} t \cdot f_X(t)\mathrm{d}t$
  - $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(t) \cdot f_X(t)\mathrm{d}t$

# Exponential RV review

- An Exponential Random Variable is defined by a parameter $\lambda$, the average number of incidents in a unit of time
  - The value of the RV is the time until the next incident
- $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
- $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$

# Review

- **Closure of the Normal Distribution**: Let $X \sim N(\mu, \sigma^2)$. Then, $aX + b \sim N(a\mu + b, a^2\sigma^2)$ i.e. linear transformations of normal random variables are still normal!
- **"Reproductive" Property of Normals**: Let $X_1, \ldots, X_n$ be independent normal random variables with $E[X_i] = \mu_i$ and $Var(X_i) = \sigma_i^2$. Let $a_1, \ldots, a_n \in \mathbb{R}, b \in \mathbb{R}$. Then

$$X = \sum_{i=1}^{n} (a_i X_i + b) \sim N\left(\sum_{i=1}^{n} (a_i \mu_i + b), \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

# Review (cont.)

- **Law of Total Probability (continuous ver.)**: $A$ is an event, and $X$ is a continuous random variable with pdf $f_X(x)$:

$$P(A) = \int_{-\infty}^{\infty} P(A|X = x) f_X(x) dx$$

- **Central Limit Theorem**: Let $X_1, \dots, X_n$ be iid random variables with $E[X_i] = \mu, Var(X_i) = \sigma^2$. Let $X = \sum_{i=1}^{n} X_i$, which has $E[X] = n\mu$ and $Var(X) = n\sigma^2$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, which has $E[X] = \mu$ and $Var(X) = \frac{\sigma^2}{n}$. $\bar{X}$ is called the *sample mean*. Then, as $n \to \infty$, $\bar{X}$ approaches the normal distribution $N(\mu, \frac{\sigma^2}{n})$. Standardizing, this is equivalent to $Y = \frac{\bar{X}-u}{\sigma/\sqrt{n}}$ approaching $N(0, 1)$. Similarly, $X$ approaches $N(n\mu, n\sigma^2)$ and $Y' = \frac{X-nu}{\sigma\sqrt{n}}$ approaches $N(0, 1)$.
  - Basically, the CLT says that sums of iid random variables, **regardless of their distribution**, tend towards a normal distribution as you add more of them together

# Review (cont.)

- **Multivariate: Discrete to Continuous:**

| | **Discrete** | **Continuous** |
|---|---|---|
| **Joint PMF/PDF** | $p_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y)$ | $f_{X,Y}(x,y) \neq \mathbb{P}(X=x, Y=y)$ |
| **Joint range/support** $\Omega_{X,Y}$ | $\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$ | $\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$ |
| **Joint CDF** | $F_{X,Y}(x,y) = \sum_{t \le x, s \le y} p_{X,Y}(t,s)$ | $F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$ |
| **Normalization** | $\sum_{x,y} p_{X,Y}(x,y) = 1$ | $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$ |
| **Marginal PMF/PDF** | $p_X(x) = \sum_y p_{X,Y}(x,y)$ | $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$ |
| **Expectation** | $\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$ | $\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$ |
| **Independence** must have | $\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$ <br> $\Omega_{X,Y} = \Omega_X \times \Omega_Y$ | $\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$ <br> $\Omega_{X,Y} = \Omega_X \times \Omega_Y$ |

# Continuity Correction

- To estimate probability that discrete RV lands in (integer) interval $\{a, \ldots, b\}$, compute probability continuous approximation lands in interval $[a - ½, b + ½]$

# Problem 1 - Will the battery last?

# Problem 1

Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with expectation 10,000 miles. If the owner wants to take a 5000 mile road trip, what is the probability that she will be able to complete the trip without replacing the battery, given that the car has already been used for 2000 miles?

# Problem 1 solution

Let $N$ be a r.v. denoting the number of miles until the battery wears out. Then $N \sim \exp(10,000^{-1})$, because N measures the ”time” (in this case miles) before an occurrence (the battery wears out) with expectation 10,000. Since this is an exponential distribution, and the expectation of an exponential distribution is $\frac{1}{\lambda}$, $\lambda = \frac{1}{10{,}000}$. Therefore, via the property of memorylessness of the exponential distribution:

$$\mathbb{P}(N \geq 5000|N \geq 2000) = \mathbb{P}(N \geq 3000) = 1 - \mathbb{P}(N \leq 3000) = 1 - \left(1 - e^{-\frac{3000}{10000}}\right) \approx 0.741$$

# Problem 2 - Normal questions

# Question 2

(a) Let X be a normal random with parameters $\mu = 10$ and $2 = 36$.
Compute $P(4 < X < 16)$.

(b) Let X be a normal random variable with mean 5. If $P(X > 9) = 0.2$, approximately what is Var(X)?

(c) Let X be a normal random variable with mean 12 and variance 4. Find the value of c such that

$$P(X > c) = 0.10.$$

# Problem 2

- (a) Let $\frac{X-10}{6} = Z$. (Recall standardization: $Z \sim N(0,1)$)

$$P(4 < X < 16) = P\left(\frac{4-10}{6} < \frac{X-10}{6} < \frac{16-10}{6}\right) = P(-1 < Z < 1) = \Phi(1) - \Phi(-1) \approx 0.68268$$

# Problem 2 Solution

- (a) Let $\frac{X-10}{6} = Z$. (Recall standardization: $Z \sim N(0,1)$)

$$P(4 < X < 16) = P\left(\frac{4-10}{6} < \frac{X-10}{6} < \frac{16-10}{6}\right) = P(-1 < Z < 1) = \Phi(1) - \Phi(-1) \approx 0.68268$$

- (b) Let $\sigma^2 = \text{\textbackslash Var}(X)$. Then

$$P(X > 9) = P\left(\frac{X-5}{\sigma} > \frac{9-5}{\sigma}\right) = 1 - \Phi\left(\frac{4}{\sigma}\right) = 0.2$$

From the phi table, we get that $\frac{4}{\sigma} = 0.845$. Therefore, $\sigma \approx 4.73 \Rightarrow \sigma^2 \approx 22.4$.

# Problem 2 Solution

- (a) Let $\frac{X-10}{6} = Z$. (Recall standardization: $Z \sim N(0,1)$)

$$P(4 < X < 16) = P\left(\frac{4-10}{6} < \frac{X-10}{6} < \frac{16-10}{6}\right) = P(-1 < Z < 1) = \Phi(1) - \Phi(-1) \approx 0.68268$$

- (b) Let $\sigma^2 = \text{\Var}(X)$. Then

$$P(X > 9) = P\left(\frac{X-5}{\sigma} > \frac{9-5}{\sigma}\right) = 1 - \Phi\left(\frac{4}{\sigma}\right) = 0.2$$

From the phi table, we get that $\frac{4}{\sigma} = 0.845$. Therefore, $\sigma \approx 4.73 \Rightarrow \sigma^2 \approx 22.4$.

- (c) Similar to (b):

$$P(X > c) = P\left(\frac{X-12}{2} > \frac{c-12}{2}\right) = 1 - \Phi\left(\frac{c-12}{2}\right) = 0.1$$

From the phi table, we get that $\frac{c-12}{2} = 1.29$. Therefore, $c \approx 14.58$

# Problem 3 – Bad Computer

# Problem 3

Each day, the probability your computer crashes is 10%, independent of every other day. Suppose we want to evaluate the computer's performance over the next $100$ days.

(a) Let $X$ be the number of crash-free days in the next $100$ days. What distribution does $X$ have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact (possibly unsimplified) expression for $\mathbb{P}(X \geq 87)$.

(b) Approximate the probability of at least 87 crash-free days out of the next 100 days using the Central Limit Theorem. Use continuity correction.

# Problem 3 solution

(a) Let $X$ be the number of crash-free days in the next 100 days. What distribution does $X$ have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact (possibly unsimplified) expression for $\mathbb{P}(X \geq 87)$.

Since $X$ counts the number of crash-free days (successes) in 100 days (trials), where each trial is a success with probability 0.9, we can see that $X$ is binomial with $n = 100$ and $p = 0.9$, or $X \sim \text{Binomial}(100, 0.9)$.

# Problem 3 solution

(a) Let $X$ be the number of crash-free days in the next 100 days. What distribution does $X$ have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact (possibly unsimplified) expression for $\mathbb{P}(X \geq 87)$.

Since $X$ counts the number of crash-free days (successes) in 100 days (trials), where each trial is a success with probability 0.9, we can see that $X$ is binomial with $n = 100$ and $p = 0.9$, or $X \sim \text{Binomial}(100, 0.9)$.

Hence, $\mathbb{E}[X] = np = 90$ and $Var(X) = np(1-p) = 9$.

# Problem 3 solution

(a) Let $X$ be the number of crash-free days in the next 100 days. What distribution does $X$ have? Identify $\mathbb{E}[X]$ and $Var(X)$ as well. Write an exact (possibly unsimplified) expression for $\mathbb{P}(X \geq 87)$.

Since $X$ counts the number of crash-free days (successes) in 100 days (trials), where each trial is a success with probability 0.9, we can see that $X$ is binomial with $n = 100$ and $p = 0.9$, or $X \sim \text{Binomial}(100, 0.9)$.

Hence, $\mathbb{E}[X] = np = 90$ and $Var(X) = np(1-p) = 9$.

$$\boxed{\mathbb{P}(X \geq 87) = \sum_{k=87}^{100} \binom{100}{k} (0.9)^k (1-0.9)^{100-k}}$$

# Problem 3 solution

(b) Approximate the probability of at least 87 crash-free days out of the next 100 days using the Central Limit Theorem. Use continuity correction.

$$\begin{aligned}
\mathbb{P}(X \geq 87) &= \mathbb{P}(86.5 < X < 100.5) = \mathbb{P}\left(\frac{86.5-90}{3} < \frac{X-90}{3} < \frac{100.5-90}{3}\right) \\
&\approx \mathbb{P}\left(-1.17 < \frac{X-90}{3} < 3.5\right) \approx \Phi(3.5) + \Phi(1.17) - 1 \approx 0.9998 + 0.8790 - 1 = 0.8788
\end{aligned}$$