CS70 - Spring 2024
Lecture 23 - April 11
Recap of Previous Lecture

- **Continuous r.v.** $X$ is described by a **prob. density function** $f: \mathbb{R} \to \mathbb{R}$ s.t.
  - $f(x) \geq 0 \ \forall x$
  - $\int_{-\infty}^{\infty} f(x) \, dx = 1$
  - $\Pr(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$

- **Note**: $f(x)$ is probability **density** per unit length
  \[ \Pr(x \leq X \leq x+dx) \approx f(x) \, dx \text{ for infinitesimal } dx \]

- **Cumulative dist. fun.** : $F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(z) \, dz$
  - $f(x) = \frac{d}{dx} F(x)$
Recap (cont.)

Continuous **uniform** distribution on \([L, R]\):

\[
f(x) = \begin{cases} 
0 & \text{if } x < L \\
\frac{1}{(R-L)} & \text{if } L \leq x \leq R \\
0 & \text{if } x > R 
\end{cases}
\]

For \(L \leq a \leq b \leq R\):

\[
\Pr\{a \leq X \leq b\} = \frac{b-a}{R-L}
\]

- **Expectation**: \(E[X] = \int_{-\infty}^{\infty} x f(x) \, dx\)
- **Variance**: \(\text{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x) \, dx - E[X]^2\)
Recap (cont.)

- **Joint distribution** of r.v.'s $X, Y$: pdf $f: \mathbb{R}^2 \to \mathbb{R}$
  
  \[ f(x,y) \geq 0 \]
  
  \[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \, dx \, dy = 1 \]

\[ \Pr[a \leq X \leq b, c \leq Y \leq d] = \int_{a}^{b} \int_{c}^{d} f(x,y) \, dx \, dy \]

- $X, Y$ are **independent** if $\forall a, b, c, d$

  \[ \Pr[a \leq X \leq b, c \leq Y \leq d] = \Pr[a \leq X \leq b] \cdot \Pr[c \leq Y \leq d] \]

$X, Y$ independent $\Rightarrow$ $f(x,y) = f_x(x) \cdot f_y(y)$
Example: Two-round game

- **Round 1**: You stake $\ell$ and win amount $X$ uniform in $[0, \ell]$
- **Round 2**: You stake $X$ and win amount $Y$ uniform in $[0, X]$

- $f(x, y) = 0$ outside red triangle
- Density of $X$ is uniform on $[0, \ell]$
- Given $X = x$, density of $Y$ is uniform on $[0, x]$
- $f(x, y) = \begin{cases} \frac{1}{\ell}x & \text{for } (x, y) \in \triangle \\ 0 & \text{otherwise} \end{cases}$
• \( f(x,y) = 0 \) outside red \( \triangle \)

• Density of \( x \) is uniform on \([0, l]\)

• Given \( x \), density of \( y \) is uniform on \([0, x]\)

\[
f(x,y) = \begin{cases} \frac{1}{l}x & \text{for } (x,y) \in \triangle \\ 0 & \text{otherwise} \end{cases}
\]

\[
f(x,y) = f_x(x) f_{y|x}(y) = \frac{1}{l} x \frac{1}{x} = \frac{1}{l}
\]

\( \text{Cf. discrete: } \Pr[X=a, Y=b] = \Pr[X=a] \Pr[Y=b \mid X=a] \)

Check: \[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \, dx \, dy = \int_{0}^{l} \int_{0}^{x} \frac{1}{lx} \, dy \, dx = 1
\]

Also: \[
E[Y] = \int_{-\infty}^{\infty} y f(x,y) \, dx \, dy = \frac{l}{4} \quad \quad \quad \quad E[X] = \int_{-\infty}^{\infty} x f(x,y) \, dx \, dy = \frac{l}{2}
\]
Today

- Exponential distribution
- Normal (Gaussian) distribution
- Central Limit Theorem
  ("averages always look like Gaussians"!)

Exponential Distribution

Continuous-time analog of Geometric distribution

Recall: \( X \sim \text{Geom}(p) \)

\[
\Pr(X = k) = (1-p)^{k-1} p
\]

Interpretation: \( X = \text{no. of trials until the first success} \) where \( p = \text{success prob.} \)

Exponential distribution measures the time we have to wait until some event happens, given that events happen at fixed rate \( \lambda \) (in continuous time)
Definition: An exponential r.v. $X$ with parameter $\lambda$ is a continuous r.v. with p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We write $X \sim \text{Exp}(\lambda)$. 

![Graph showing exponential probability density functions with different values of $\lambda$.](image)
**Definition:** An exponential r.v. $X$ with parameter $\lambda$ is a continuous r.v. with p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We write $X \sim \text{Exp}(\lambda)$

Check p.d.f. conditions:

- $f(x) \geq 0 \quad \checkmark$
- $\int_{-\infty}^{\infty} f(x) \, dx = \int_{0}^{\infty} \lambda e^{-\lambda x} \, dx$

\[
= -e^{-\lambda x} \bigg|_{0}^{\infty} = 1 \quad \checkmark
\]
Connection with Geometric distribution

\[ X \sim \text{Exp}(\lambda) \]

Then \( P(X > t) = \int_t^\infty \lambda e^{-\lambda x} \, dx = \left. -e^{-\lambda x} \right|_t^\infty = e^{-\lambda t} \)
Connection with Geometric distribution

\( X \sim \text{Exp}(\lambda) \)

Then \( \Pr(X > t) = \int_t^{\infty} \lambda e^{-\lambda x} \, dx = -e^{-\lambda x} \bigg|_t^{\infty} = e^{-\lambda t} \)

Discrete time setting: Since we perform one trial every 8 seconds, with fixed success prob. \( \lambda \) per unit time.

Then \# trials until first success is \( Y \sim \text{Geom}(p) \)

where \( p = \lambda \delta \). And time until first success is \( T = \delta Y \) (seconds).
Connection with Geometric distribution

\[ X \sim \text{Exp}(\lambda) \]

Then \( \Pr \{ X > t \} = \int_t^\infty \lambda e^{-\lambda x} \, dx = -e^{-\lambda x} \bigg|_t^\infty = e^{-\lambda t} \)

Discrete time setting: Suppose we perform one trial every \( \delta \) seconds, with fixed success prob. \( \lambda \) per unit time.

Then the number of trials until first success is \( Y \sim \text{Geom}(p) \), where \( p = \lambda \delta \). And the time until first success is \( T = \delta Y \) (seconds).

Now let \( \delta \to 0 \) (so also \( p \to 0 \)).

Then \( \Pr \{ T > t \} = \Pr \{ Y > \frac{t}{\delta} \} = (1-p)^{t/\delta} = (1-\lambda \delta)^{t/\delta} \to e^{-\lambda t} \)

as \( \delta \to 0 \) (\( t, \lambda \) fixed).
Expectation & Variance

\[ E[X] = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{\infty} x \lambda e^{-\lambda x} \, dx \]

\[
\begin{align*}
\text{Recall: } \int u \, dv &= uv - \int v \, du \\
\int_{0}^{\infty} x \lambda e^{-\lambda x} \, dx &= 0 - \frac{1}{\lambda} e^{-\lambda x} \bigg|_{0}^{\infty} = \frac{1}{\lambda}
\end{align*}
\]
**Expectation & Variance**

\[
E[X] = \frac{1}{\lambda}
\]

\[
E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{0}^{\infty} \frac{x^2 \lambda e^{-\lambda x}}{\lambda} dx
\]

\[
= -x^2 e^{-\lambda x} \bigg|_{0}^{\infty} + \int_{0}^{\infty} 2x e^{-\lambda x} dx
\]

\[
= 0 + \frac{2}{\lambda} E[X] = \frac{2}{\lambda}
\]

\[
\text{Var}(X) = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
\]

Recall: "Integration by Parts":
\[
\int u dv = uv - \int v du
\]
Normal (a.k.a. Gaussian) Distribution

For any $M \in \mathbb{R}$ and $\sigma > 0$, a continuous r.v. $X$ with p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-M)^2}{2\sigma^2}}$$

is a normal r.v. with parameters $(M, \sigma^2)$. We write $X \sim N(M, \sigma^2)$.

$M=0, \sigma=1 \rightarrow$ standard normal distribution
\[ f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

**Check:**
- \( f(x) \geq 0 \)
- \( \int_{-\infty}^{\infty} f(x) \, dx = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = 1 \)
Fact: All normal distributions are the same up to shifts and scaling.

If \( X \sim N(\mu, \sigma^2) \) then \( Y = \frac{X - \mu}{\sigma} \sim N(0, 1) \)

Proof: \( \Pr[a \leq Y \leq b] = \Pr[\sigma a + \mu \leq X \leq \sigma b + \mu] \)

\[
= \frac{1}{\sqrt{2\pi} \sigma^2} \int_{\sigma a + \mu}^{\sigma b + \mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx
\]

Change of variable: \( y = \frac{x - \mu}{\sigma} \)

\[
= \frac{1}{\sqrt{2\pi}} \int_{\sigma a + \mu}^{\sigma b + \mu} e^{-\frac{y^2}{2}} \, \sigma \, dy
\]

\[
= \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{y^2}{2}} \, dy
\]
Fact: All normal distributions are the same up to shifts and scaling.

If $X \sim N(\mu, \sigma^2)$ then $Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$

Books/internet usually just give c.d.f. of standard normal, i.e., you can look up $Pr[Y \leq c]$ for $Y \sim N(0, 1)$

But then if $X \sim N(\mu, \sigma^2)$ you can get

$$Pr[X \leq c] = Pr\left[Y \leq \frac{c - \mu}{\sigma}\right]$$

from table
Expectation & Variance

Suppose $X \sim N(0, 1)$ is standard normal.

Then p.d.f. is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$E[X] = \int_{-\infty}^{\infty} x f(x) \, dx = \frac{1}{\sqrt{2\pi}} \left[ \int_{-\infty}^{0} x e^{-x^2/2} \, dx + \int_{0}^{\infty} x e^{-x^2/2} \, dx \right] = 0$$

$e^{-x^2/2}$ symmetric about 0.
Expectation & Variance

Suppose \( X \sim N(0, 1) \) is standard normal

Then p.d.f. is \( f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \)

\[
E[X] = \int_{-\infty}^{\infty} x f(x) \, dx = \frac{1}{\sqrt{2\pi}} \left[ \int_{-\infty}^{0} x e^{-x^2/2} \, dx + \int_{0}^{\infty} x e^{-x^2/2} \, dx \right] = 0
\]

\( e^{-x^2/2} \) symmetric about 0

\[
E[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \, dx
\]

\[
= \frac{1}{\sqrt{2\pi}} \left[ -x e^{-x^2/2} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2} \, dx \]

\[
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \, dx = 1 \quad \Rightarrow \quad \text{Var}(X) = 1
\]

\( u = x \quad dv = xe^{-x^2/2} \, dx \\
\quad \quad \quad du = dx \quad v = -e^{-x^2/2} \)
Expectation & Variance (cont.)

For $X \sim N(0, 1)$: $E[X] = 0$ \hspace{1cm} \text{Var}(X) = 1$

For $X \sim N(\mu, \sigma^2)$, then $Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$ so:

$E[Y] = E\left[\frac{X - \mu}{\sigma}\right] = 0$

$\Rightarrow E[X] = \mu$

$\text{Var}[Y] = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = 1$

$\Rightarrow \text{Var}(X) = \sigma^2$

This explains notation $\mu, \sigma^2$
**Sum of Independent Normals**

**Fact**: If $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent, and $a, b$ are constants, then

$$aX + by \sim N(0, a^2 + b^2)$$

**Note**: Expectation & variance are obvious from:

$$E[aX + by] = aE[X] + bE[Y] = 0$$

$$\text{Var}(aX + by) = \text{Var}(ax) + \text{Var}(by) \quad \text{[independent \, \%]}$$

$$= a^2 \text{Var}(X) + b^2 \text{Var}(Y) = a^2 + b^2$$

Not so obvious: $aX + by$ is Normal
Proof: Since $X, Y$ are independent their joint p.d.f. is

$$f(x, y) = f_x(x)f_y(y) = \frac{1}{2\pi} e^{-\left(\frac{x^2+y^2}{2}\right)}$$

This function is rotationally symmetric around 0.

So $\alpha X + \beta Y = t$ is a vertical "slice" that can be rotated to $X = \frac{t}{\sqrt{a^2+b^2}}$ (preserves distance from 0).

Details: Note 21
Fact: If \( X \sim N(0,1) \) and \( Y \sim N(0,1) \) are independent, and \( a, b \) are constants, then
\[
aX + bY \sim N(0, a^2 + b^2)
\]

Generalization: If \( X \sim N(M_x, \sigma_x^2) \) and \( Y \sim N(M_y, \sigma_y^2) \) are independent, then
\[
aX + bY \sim N(aM_x + bM_y, a^2\sigma_x^2 + b^2\sigma_y^2)
\]

Proof: Apply above Fact to standard normals
\[
\frac{X - M_x}{\sigma_x} \quad \text{and} \quad \frac{Y - M_y}{\sigma_y}
\]
Central Limit Theorem

Recall: average of many independent samples of the same r.v.

\[ X_1, X_2, \ldots \text{ i.i.d.} \quad E[X_i] = \mu, \quad \text{Var}(X_i) = \sigma^2 \]

Sample average: \[ \frac{1}{N} S_N \quad \text{where } S_N = X_1 + \ldots + X_N \]

Amazing Fact: As \( N \to \infty \), the distribution of \[ \frac{1}{N} S_N \] approaches \( N(\mu, \sigma^2/N) \)

\[ E\left[ \frac{1}{N} S_N \right] = \frac{1}{N} \sum E[X_i] = \mu \]
\[ \text{Var}\left( \frac{1}{N} S_N \right) = \frac{1}{N^2} \sum \text{Var}(X_i) = \sigma^2/N \]
Scale $\frac{1}{N}S_n$ so that limit is standard normal:

$$\frac{\frac{1}{N}S_n - \mu}{\sigma/\sqrt{N}} = \frac{S_n - N\mu}{\sigma\sqrt{N}}$$

Central Limit Theorem: For i.i.d. r.v.'s $X_i$ with $\mathbb{E}[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$, the distribution of $\frac{S_n - N\mu}{\sigma\sqrt{N}}$ converges to $N(0,1)$ as $N \to \infty$.

I.e.: $\Pr \left[ \frac{S_n - N\mu}{\sigma\sqrt{N}} \leq c \right] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-x^2/2} \, dx$ as $N \to \infty$

for any constant $c$. 
**CLT: Example**

\[ X_i \sim \text{Geom}(1/6) \]

\[ \frac{1}{N} (X_1 + \ldots + X_N) \]

\[ M = \frac{1}{p} = 6, \quad \sigma^2 = \frac{1-p}{p^2} = 30 \]

\[ \text{Distribution of } S_{N/N} \text{ for } N=1 \]

\[ \text{Distribution of } S_{N/N} \text{ for } N=2 \]

\[ \text{Distribution of } S_{N/N} \text{ for } N=5 \]

\[ \text{Distribution of } S_{N/N} \text{ for } N=10 \]

\[ \text{Distribution of } S_{N/N} \text{ for } N=20 \]

\[ \text{Distribution of } S_{N/N} \text{ for } N=100 \]

"Width" of distribution \( \approx \frac{\sigma}{\sqrt{N}} \rightarrow 0 \text{ as } N \rightarrow \infty \)