**הסתברות ותהליכים סטוכסטיים (פרופ' יזהר בר-גד)**

**12/05/2020, 17/05/2020**

**תורת המידע**

*תורת האינפורמציה היא ענף במתמטיקה שייסד קלוד שאנון בשנות ה-40. הקטליזטור המרכזי להתפתחות התורה היה העיסוק באיסוף מודיעין במסגרת מלחמת העולם השנייה והוא עסק בשאלות איזה מידע ניתן ואיזה מידע לא ניתן להעביר דרך מערכת. תורת המידע מציעה דרכים למדידות כמותיות של מידע ושל היכולת של מגוון מערכות לשדר, לאחסן ולבצע פעולות שונות על מידע. התורה משמשת במגוון תחומים בהם תקשורת, כיווץ מידע, קריפטוגרפיה, מדעי המחשב, ביולוגיה, פסיכולוגיה, מדעי המוח ועוד. דוגמא בולטת לשימוש בתורת האינפורמציה- כשאנחנו מצלמים, יש לנו מצלמה עם 20 מיליון פיקסלים, ועל כל פיקסל ניתן לשמור 8 ביטים של R, G ו-B, כלומר בחישוב פשוט מתקבל שבכל תמונה צריך לשמור 60 מיליון בתים (כל בית הוא 8 ביטים), אולם בפועל אנחנו לא שומרים בכל תמונה את כל המידע אלא שומרים פחות כדי לחסוך מידע. כך גם למשל בתקשורת טלפונית- קול עובר כיווץ ומגיע לצד השני. כל זאת הודות לתורת המידע כפי שפותחה בשנות ה-40.*

*שני ספרים מומלצים בתחום:*

*Elements of Information Theory, T. Cover & J. Thomas*

*Information Theory, Inference, and Learning Algorithms, David J.C. MacKey.*

*את התחום בתורת האינפורמציה המשיכו לפתח תלמידיו שלקחו זאת לכיוונים של פיזיקה או לכיוון ההפוך של סטטיסטיקה ומדעי המחשב. מסיבה זו יש מונחים רבים כפולים בתחום, כאשר אנחנו נעבוד בעיקר עם המינוחים של הסטטיסטיקאים.*

*אחד המאפיינים הבסיסיים של מערכת מידע נקרא* ***אנטרופיה*** *והוא מייצג את מידת חוסר הוודאות לגבי המצב של אותה מערכת. האנטרופיה נמדדת ע"י מספר הביטים הנדרשים על מנת לתאר באופן מלא את המצב של המערכת. נסתמך על כך שסמלים אחרים יכולים לעבור טרנספורמציה בקלות לביטים, למשל כל האותיות באנגלית יכולות להיות מתוארות בעזרת רצפים של 5 ביטים. כמו כן, האנטרופיה נחשבת למספר השאלות של כן ולא הנדרשות על מנת לאפשר הבנה מלאה של המערכת. הסוג הזה של האנטרופיה נקרא "האנטרופיה של שאנון" או "אנטרופיה של מידע" במטרה להבחין בינה לבין האנטרופיה שיש בתרמודינמיקה.*

***דוגמא להמחשה: הטלת מטבע***

*מטילים מטבע ומקבלים עץ או פלי, כאשר מסמנים בעזרת ביטים את התוצאות- 0 לעץ, 1 לפלי. על בסיס צורת הקידוד הזו, סדרת הטלות המטבע- עץ-עץ-פלי-עץ-פלי היא אקוויוואלנטית לסדרת הביטים 00101. נדרש בדיוק ביט אחד כדי לייצג כל הטלה. כל עוד אנחנו יודעים את הטרנספורמציה, אנחנו יכולים לדעת את הקידוד הלוך והקידוד חזור לכל סדרת ביטים.*

*נניח שאנחנו מטילים שני מטבעות באופן סימולטני, אנחנו יכולים לקודד את התוצאות באופן הבא:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *מטבע א'* | *עץ* | *עץ* | *פלי* | *פלי* |
| *מטבע ב'* | *עץ* | *פלי* | *עץ* | *פלי* |
| *קידוד* | *00* | *01* | *10* | *11* |

*על בסיס סכמת הקידוד הזו הרצפים הבאים של הטלות מטבע אקוויוואלנטיים:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *חזרה מס'* | *1* | *2* | *3* | *4* |
| *מטבע א'* | *עץ* | *פלי* | *פלי* | *פלי* |
| *מטבע ב'* | *עץ* | *עץ* | *פלי* | *עץ* |

*מקודד ל-00101110*

*כלומר נדרשים בדיוק 2 ביטים להביע כל הטלה סימולטנית. מעניין אותנו האם אפשר להביע זאת בפחות ביטים מכיוון שאם הם הוגנים יש 4 תוצאות אפשריות לכל ניסוי והמספר המינימלי להביע 4 תוצאות אפשריות כולל שני ביטים. אולם, אם המטבעות לא הוגנים, ויש משהו שחוזר על עצמו, ניתן להשתמש בחזרתיות כדי "לחסוך" ביטים.*

*כעת נניח שלא משנה לנו סדר ההטלה, איך זה משפיע? כלומר מעניין אותנו רק אם קיבלנו זוג מתואם (שני עצים או שני פלי) או זוג בלתי מתואם (עץ ופלי או פלי ועץ). נוכל לייצג את ההסתברות לכל זוג- שני עצים 25%, שני פלי 25% וזוג מעורב 50%. נוכל להשתמש בסכמת קידוד שונה- זוג מעורב 0, זוג עצים- 10 וזוג פלי- 11. במצב כזה סכמת הקידוד תביא לאקוויוואלנטיות אחרת:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *חזרה מס'* | *1* | *2* | *3* | *4* |
| *מטבע א'* | *עץ* | *פלי* | *פלי* | *פלי* |
| *מטבע ב'* | *עץ* | *עץ* | *פלי* | *עץ* |

*מקודד ע"י 100110*

*איך יודעים לחזור חזרה? כל פעם שמקבלים 0 יודעים שזה מעורב וכל פעם שמקבלים 1 יודעים שצריך לקרוא גם את הביט הבא כדי לקרוא את התוצאה- שני עצים או שני פלי. כל עוד קראנו את הקוד מהמקום הנכון, התחלנו לקרוא אותו היכן שהוא מתחיל, נוכל תמיד לשחזר חזרה את ההודעה.*

*המספר הממוצע של ביטים שאנחנו משתמשים משתנה שכן ב-25% מהמקרים אנחנו משתמשים ב-2 ביטים (עבור שני עץ), ב-25% נוספים שוב משתמשים ב-2 ביטים (עבור שני פלי) וב-50% מהמקרים בביט אחד. לכן הממוצע הוא:*

*מה המשמעות של 1.5 ביט? משמעותו רק ייצוג של ממוצע לגבי מידע- לעיתים יתקבל ביט 1 ואז לא יהיה מספיק מידע ולעיתים 2 ביטים ואז יהיה עודף. השאלה היא רלוונטית בעצם רק למסר בעל אורך גבוה- למשל עבור מיליון תוצאות נצטרך מיליון וחצי ביטים. אנחנו לא ננסה לשאול את השאלה ההסתברותית עבור מיליון הטלות האם צריך בהכרח מיליון וחצי ביטים, אבל נתבסס על כך שבממוצע אפשר להשתמש במיליון וחצי ואנחנו מספיק קרובים לממוצע. בספרים המוצעים מעלה יש גם עיסוק בשאלות האלה- כמה אחוזים מהעברת המידע תהיה ב-worst case (למשל יידרשו 2,000,000 ביטים) וכמה אחוזים בממוצע- כלומר איך מתפלגת העברת האינפורמציה.*

***אנטרופיה ואינפורמציה***

*כזכור, אנטרופיה של מערכת היא מידת חוסר הוודאות לגבי המצב שלה, כלומר המספר המצופה של ביטים הנדרשים לתאר באופן מלא את המצב של המערכת. למשל בדוגמא האחרונה של הטלות המטבע, נדרשו לנו 1.5 ביטים על מנת לתאר את התוצאה. ניתן גם להוכיח כי זהו המצב המינימלי של האנטרופיה עבור מצב זה.*

*כעת ניתן להגדיר מהו אינפורמציה- אינפורמציה היא הסכום שבו חוסר הוודאות שלנו יורדת בו בהינתן ידע חדש. בדוגמא של הטלת מטבע של שני מטבעות מסונכרנים, אם קיבלנו ידע חדש שלא אומר לנו מה התוצאה אלא רק ששני המטבעות היו זהים (מסומן ב-1) אנחנו מקבלים 0.5 ביטים של מידע (ונותר רק ביט אחד של חוסר וודאות). לכן אפשר להגדיר אינפורמציה בעזרת דלתא- הפער שבין המידע שלפני לבין המידע שאחרי.*

*אנטרופיה היא האורך המצופה של ביטים הנושאים הודעה בינארית המעבירה מידע. יש לה גם שמות נוספים כגון: הסיבוכיות של הקוד, חוסר וודאות, אינפורמציה חסרה/נדרשת, הפתעה צפויה, תוכן האינפורמציה של המערכת ועוד. מבחינה היסטורית, האנטרופיה מוגדרת בתרמודינמיקה קלאסית בתור כמות החום הבלתי שמיש שיש במערכת ובתרמודינמיקה סטטיסטית מכונה מידת חוסר הסדר של המערכת, כאשר השניים מוכחים כאקוויוואלנטיים זה לזה. בספרים שנכתבו ע"י פיזיקאים לרוב משתמשים בקשר האקוויוואלנטי הזה בדיון על אנטרופיה.*

***אינפורמציה לפי שאנון***

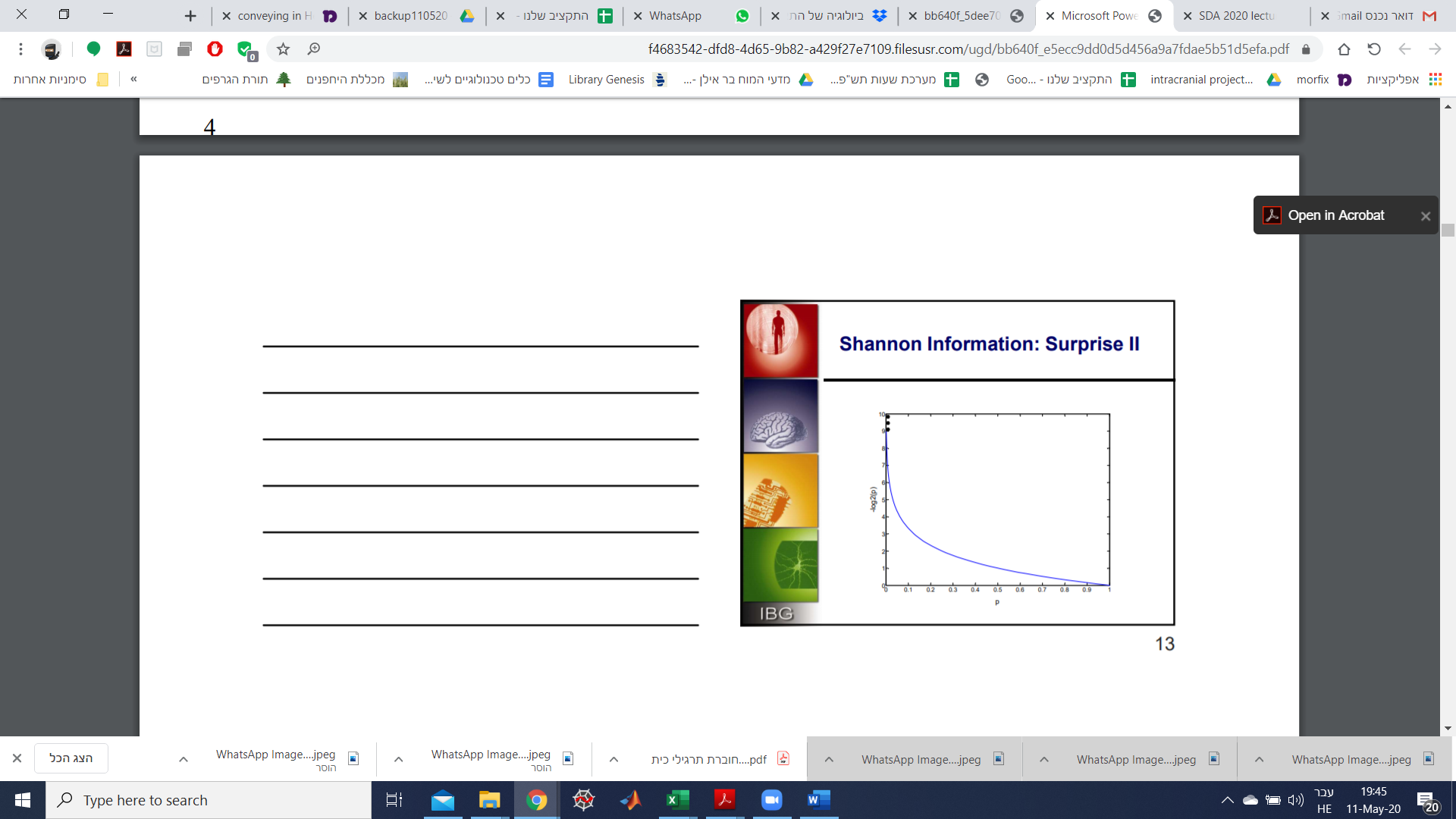
*היחידה הקטנה ביותר של מידע נקראת "ביט". ביט אחד הוא כמות המידע הנדרש כדי לבחור בין שתי תוצאות בעלות סבירות שווה (לדוגמא הטלת מטבע). בין המאפיינים של האינפורמציה הם היותה בלתי תלויה באירועים מתווספים והיותה שווה לאפס אם אנחנו כבר יודעים את התוצאה. כך חיפש שאנון איזו פונקציה שומרת את התכונות הללו ופיתח את הפונקציה של "הפתעה".*

*באופן אינטואיטיבי ככל שאירוע נדיר יותר כך ההפתעה שלנו ממנו יותר גדולה. ההפתעה של אירוע אחד היא גבוהה עבור אירועים בלתי צפויים (בעלי הסתברות נמוכה) ונמוכה עבור אירועים צפויים. כלומר:*

*אירועים בלתי תלויים מקיימים ונקבע עבור הפונקציה של ההפתעה:*

*זאת מכיוון שאנחנו רוצים שמדד ההפתעה יסתכם עבור אירועים בלתי תלויים ולא יהיה כפול. הפונקציה שעונה לכל המאפיינים הללו היא:*

*הבחירה בבסיס 2 נובעת מהרצון להשתמש בביטים, אך ניתן להשתמש בבסיסים שונים- למשל בסיס עשרוני יביא לשימוש בקידוד ע"י ספרות, בסיס טבעי יביא לשימוש בקידוד ע"י natural bits- nits ועוד. הסיבה לכך שיש מינוס לפני ה- היא שאנחנו רוצים להגדיר בצורה חיובית ולא שלילית את מידת ההפתעה (כלומר ישאף ל- בקרבת ה-0 ולא ל-).*



*נשים לב לנוסחאות השימושיות עבור לוגריתמים:*

***הגדרת האנטרופיה:***

*עד כה דיברנו רק על רעיונות להבנת משמעות האנטרופיה. אולם האנטרופיה היא הערך הממוצע של הפתעה עבור כל התצפיות האפשריות:*

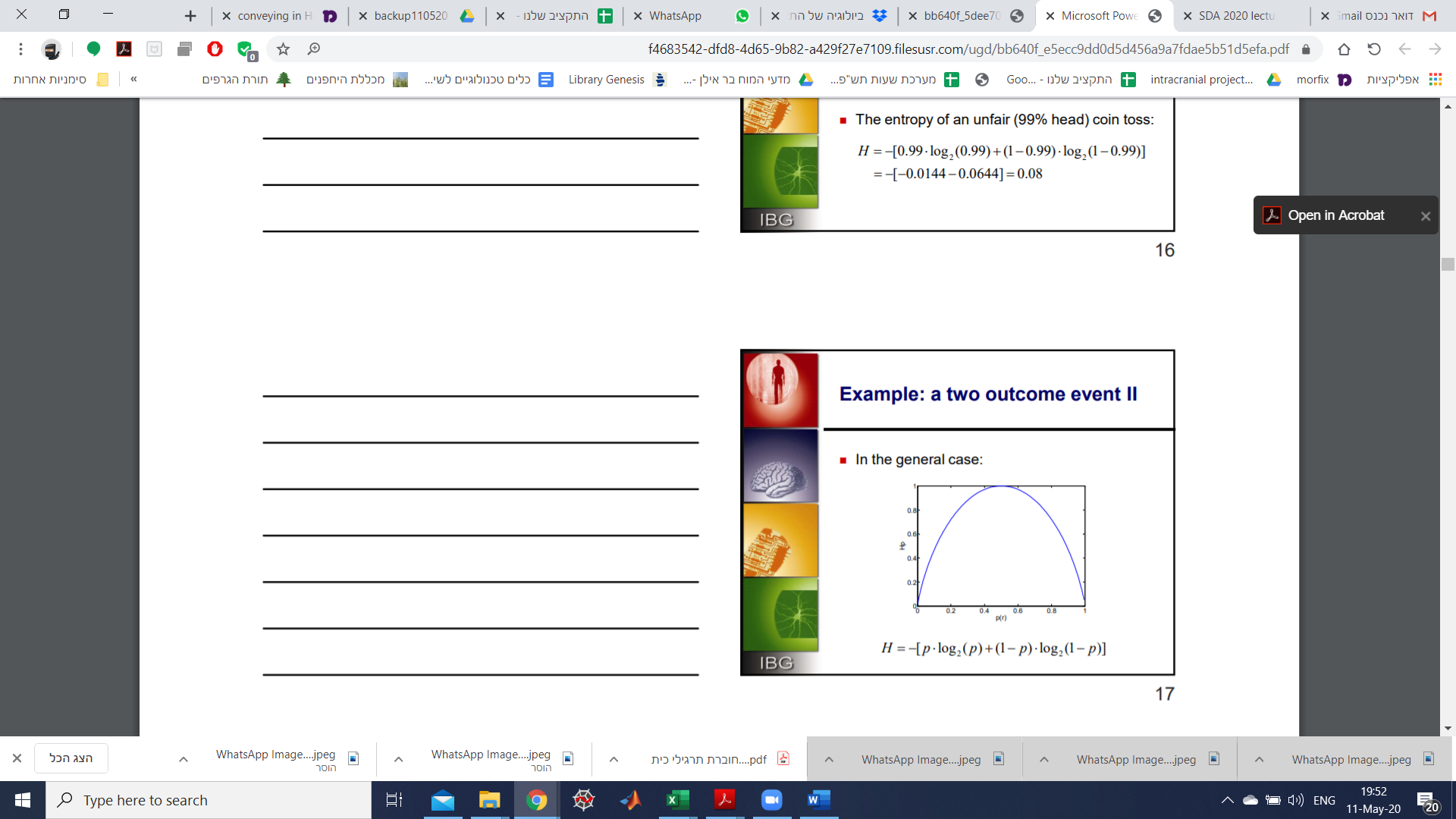
*במקרה הדיסקרטי:*

*ההגדרה הזו פותרת את הבעיה של התבדרות פונקציית ההפתעה סביב ה- מכיוון שהיא כופלת בהסתברות ולכן כאשר האיבר יתאפס.*

*נדגים את חישוב האנטרופיה עבור אירוע שיש לו שתי תוצאות אפשריות- למשל הטלת מטבע הוגן:*

*תוצאה זו מתאימה בדיוק ל-מה שהצגנו בדוגמא עם הטלת מטבע יחיד. עבור מטבע שאיננו הוגן (עם הטייה של 99 אחוזים לעץ):*

*כלומר אם נבצע ניסוי עם מיליון חזרות של הטלת מטבע- במטבע הוגן יידרשו לנו מיליון ביטים ובמטבע הלא הוגן 80,000- זהו החסם התחתון להעברת אינפורמציה ואת ההוכחה נראה בהמשך. אם ההסתברות היא , מטבע לא הוגן כלל, יביא לאנטרופיה של 0. אם ניקח את המקרה הכללי:*



***מאפיינים נוספים של האנטרופיה:***

* *האנטרופיה תמיד חיובית: הנוסחא שלה כוללת איברים חסומים ב-1 ולכן היא תמיד חיובית.*
* *המקסימום של הוא קבוע, עבור הוא יהיה והוא קורה כאשר כל התצפיות האפשריות מתפלגות בצורה אחידה.*
* *המינימום של הוא פונקציית דלתא שווה ל-0.*
* *ככל שהאנטרופיה גבוהה יותר, לומדים יותר (בממוצע) ע"י התבוננות בערכים של המשתנה המקרי.*
* *ככל שהאנטרופיה יותר גבוהה, יכולים החיזוי של הערכים של המשתנה המקרי יורדת.*

*נתבונן בדוגמא נוספת פשוטה לחישוב אנטרופיה- אם כל התוצאות האפשריות של סיטואציה בעלי הסתברות שווה, אז מידת חוסר הוודאות שלנו לגבי איזה אחת מהתוצאות תקרה ניתנת לחישוב ע"י:*

*למשל אם יש לנו 8 מטבעות שאחד מהם מזוייף, וידוע שכל האחרים אמיתיים ולאותו אחד מזוייף יש משקל שונה מהאחרים האנטרופיה נותנת לנו את חישוב המספר המינימלי שצריך לשקול מאזניים מאוזנים על מנת לקבוע איזה אחד מהם מזוייף. האנטרופיה יכולה להשתנות כתלות במספר- למשל אם יש שבעה מטבעות שאחד מהם מזוייף, או אם יש תשעה מטבעות.*

*תהליך קידוד המבוסס על אנטרופיה יכול למשל כך במקרה שיש לנו 4 סמלים (A,C,G,T), כמו למשל ב-DNA. ניתן לשאול את השאלה כמה מידע יש לכל היותר בגנום האנושי. נניח כי כל אחד מהם קורה בהסתברויות שונות:*

*הם מובילים לחישוב ההפתעות הבאים:*

*על כן הממוצע של חוסר הוודאות נתון ע"י האנטרופיה:*

*מהבחינה הזו המשמעות היא שאם רוצים לשלוח מידע על אדם נוכל לשלוח בממוצע על כלל האנשים שלכל אחד מהם יש מיליון גנים, צריך 1,750,000 ביטים בממוצע לכל אדם. כיצד מייצרים קודים כאלה? לא נעסוק בשאלות, אלא נתבונן בכמה דוגמאות.*

*האפשרות הפשוטה היא הרי קידוד בן 2 ביטים לכל סמל:*

*אפשרות אחרת היא לקודד לפי מידת ההפתעה והיא יוצאת הטובה ביותר:*

*קידוד של השרשרת ACATGAAC שבה ההסתברויות מתאימות בדיוק להסתברות המופיעה מעלה מקודד בכל אחת מהשיטות באופן הבא-*

1. *שיטה ראשונה 0001001110000001- 16 ביטים בסה"כ, 2 ביטים לכל סמל*
2. *שיטה שנייה- 10110010001101- 14 ביטים בסה"כ (1.75 ביטים לכל סמל)*

*אם ההסתברויות אינן חזקות שלמות של 2 יש תחומים שלמים שעוסקים בתהליך הקידוד, ובאופן כללי עוסק בשאלות האם ניתן במקרה הזה למצוא אפילו דרך קצרה יותר לקידוד? ובמקרה הכללי, האם אפשר לתת נוסחא כללית לקידוד האופטימלי? שאלות אלה הן בדיוק השאלות המטופלות תחת הנושא של דחיסת מידע, וניתן לקרוא עוד על כך בספר Elements of Information Theory, T. Cover & J. Thomas, Chapter 5.*

***אנטרופיה משותפת***

*אם נתבונן באנטרופיה המשותפת לשני משתנים, נוכל להגדירה כוקטור יחיד עם משתנה רנדומלי:*

*ובמקרה הדיסקרטי:*

*נשים לב כי במקרה של התפלגויות בלתי תלויות מתקבל שהאנטרופיה המשותפת היא סכום האנטרופיות.*

*נוכל גם להגדיר אנטרופיה מותנית- אותה נוסחא, אך תוך שימוש בצפיפות מותנית:*

*במילים אחרות אנחנו ממצעים סביב ההסתברות המשותפת של שני ערכים (ולא סביב ההסתברות המותנית) ורק ב-log מופיעה ההסתברות המותנית. משמעותה של האנטרופיה המותנית- בהינתן שאנחנו יודעים את X עד כמה אנחנו לא יודעים את Y. אם למשל יש לנו סדרה של שתי הטלות מטבע- השני תמיד נופל כמו הראשון, מספיק לדעת את כל ההטלות של X כדי לדעת את כל ההטלות של Y כי השני תמיד זהה לראשון. במקרה הקיצון השני שבו שתי ההטלות בלתי תלויות אז נקבל כי בחישוב האינטואיטיבי (שיוכח להלן ככל השרשרת):*

*כלומר כל מידע על הטלת לא מסייע בקבלת מידע על הטלת .*

*כלל השרשרת של האנטרופיה הוא:*

*הוכחה:*

*לפיכך:*

*זוהי המקבילה האנטרופית למשוואה של כלל בייס, שבה עברנו לעולם הלוגריתמי ולכן עברנו ממכפלות וחלוקות לחיבור וחיסור. כעת נוכל להשתמש בכלי של האנטרופיה המותנית כדי להשתמש בכלי חשוב יותר של אינפורמציה מתואמת.*

***אינפורמציה מתואמת***

*האנטרופיה אומרת לנו כמה אנחנו יכולים ללמוד (מילים אחרות לכמה אנחנו לא יודעים). המידע המתואם בין ל- מוגדר ע"י השאלות:*

* *כמה אנחנו לומדים על על ידי התבוננות ב-?*
* *כמה אנחנו יודעים יותר על לאחר התבוננות על ?*
* *כמה קל לחזות את לאחר התבוננות ב-?*

*לפיכך אפשר גם לתרגם את כל השאלות הללו עד כמה האנטרופיה של יורדת כתוצאה מהתבוננות ב-? זוהי בדיוק ההגדרה של מידע מתואם:*

*למשל כמות המידע שיש לנו לפני שאנחנו מקבלים גירוי וכמות המידע שיש לנו לאחר שאנחנו מקבלים גירוי בדיוק מתוארת ע"י המשוואה לעיל. מכאן נוכל להגדיר את האינפורמציה המתואמת:*

*שני השוויונות האחרונים הם השמישים ביותר- שני תהליכי מיצוע המביעים את שני התהליכים והקשר ביניהם תוך התבססות על כלל בייס. אפשר לראות שמעצם הגדרת הנוסחא והשוויון האחרון , שכן . נראה מיד תכונות נוספות פרט לסימטריה של אינפורמציה מתואמת.*

*כזכור משיעור ההשלמה העצמי של בייס, נחזור לדוגמא של הרופא שצריך להבחין בין שפעת שלה יש הסתברות לבין דלקת חמורה בהסתברות . יש לו שני מבחנים והפעם נסתכל עליהם דרך המשקפיים של כמה אינפורמציה נותנת כל בדיקה.*

*נשים לב כי השורות לא מסתכמות ל-1 ולכן זוהי איננה הסתברות משותפת (joint probability) אלא הסתברות מותנית (conditional probability) ולכן מה שלפנינו זה הסתברות מותנית. מכיוון שרק הטורים מסתכמים ל-1 הרי שההסתברות המופיעה כאן היא :*

|  |  |  |
| --- | --- | --- |
| *בדיקת דם* | *שפעת* | *דלקת* |
| *חיובי* | *0.2* | *0.7* |
| *שלילי* | *0.8* | *0.3* |

|  |  |  |
| --- | --- | --- |
| *בדיקת שתן* | *שפעת* | *דלקת* |
| *חיובי* | *0.1* | *0.5* |
| *שלילי* | *0.9* | *0.5* |

*איזה אחד מהמבחנים נותן יותר מידע על המצב של המטופל?*

*קיבלנו את המידע על ההסתברות המותנית ונקבל:*

*לכן, בדיקת הדם יותר אינפורמטיבית. כדי לראות האם האנטרופיה יורדת אפשר לבדוק גם ועדיין נוכל לראות שגם במצב שבו שתי הבדיקות בלתי-תלויות לחלוטין עדיין לא יירד משמעותי כי אם הם בלתי-תלויים (שזה מצב המקסימום) אז יתקיים ולכן אפילו במצב הזה נקטין את האנטרופיה לכל היותר ל- וזה עדיין אנטרופיה יחסית גדולה.*

***מאפיינים של אינפורמציה מתואמת***

*אם ו- בלתי תלויים אחד בשני, אז האינפורמציה המתואמת שווה לאפס:*

*היא תמיד קטנה יותר מהאנטרופיה:*

*לא ניתן להגדיל אותה ע"י פונקציה מתמטית:*

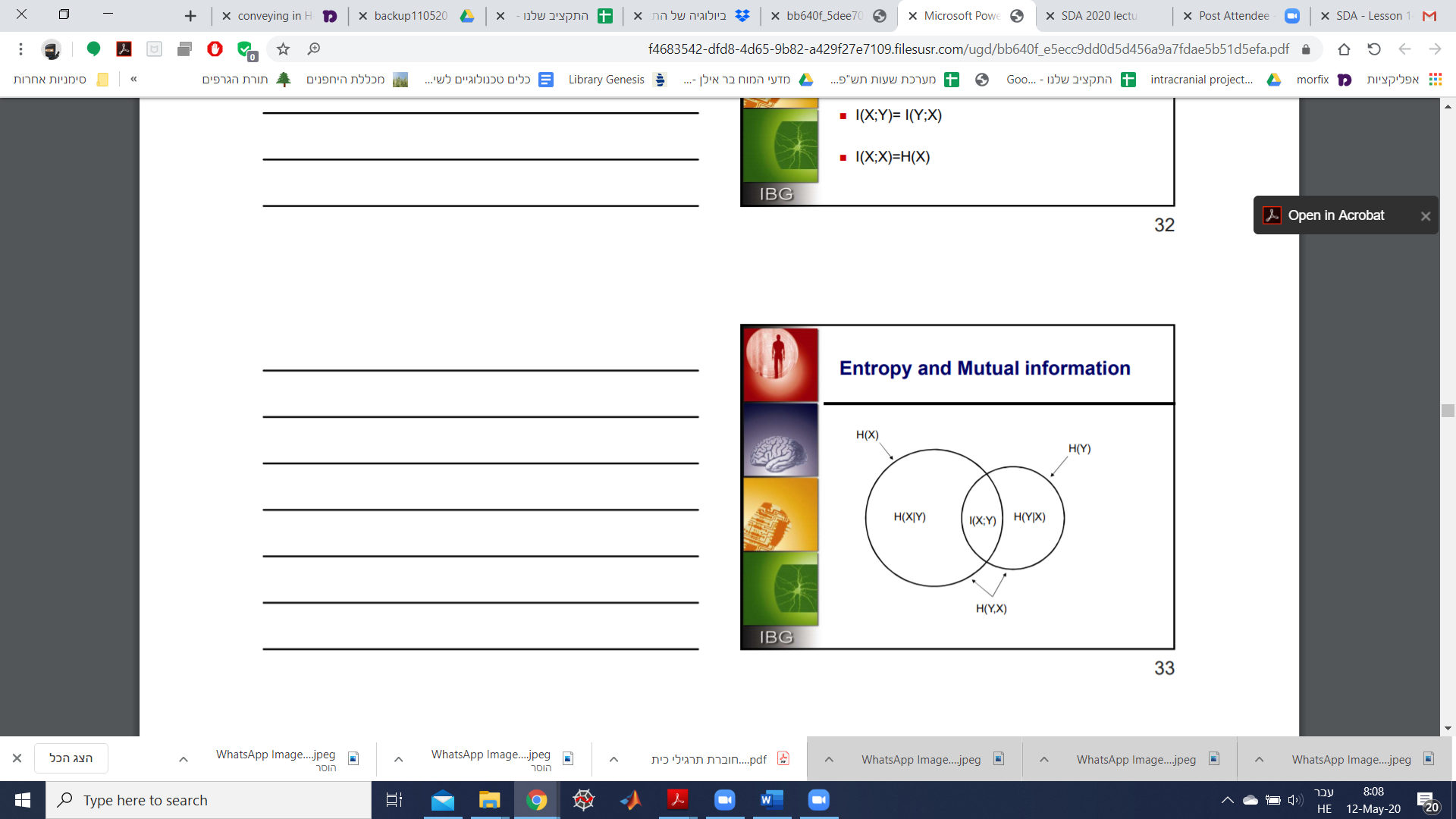
*במילים אחרות- אין כל דרך, לא דרך רשתות עצביות, או כל תהליך חישובי אחר שיאפשר להכניס יותר אינפורמציה.*

*כמה מאפיינים נוספים-*

*כזכור ההגדרה היא-*

*מהסימטריות שקיבלנו-*

*לסיכום, נוח לזכור את הקשר בין אנטרופיה לאינפורמציה מתואמת בעזרת הקשר הבא. במצב שהם בלתי תלויים אין חפיפה בין המעגלים, ואפשר לזכור שאם עושים איחוד בין ו- צריך להוריד את החיתוך פעם אחת כי הוא נספר פעמיים:*



***אנטרופיה יחסית- התבדרות קולבק-ליבר (KL)***

*התבדרות קולבק-ליבר היא ה'מרחק' (לא אוקלידי) הנמדד בין התפלגויות של הסתברויות. המרחק הזה איננו סימטרי כך שהמרחק בין שתי התפלגויות שאינו שווה למרחק בין ולכן קוראים להפרש . במקרה שלנו המרחק בין שתי התפלגויות מוגדר ע"י התבדרות KL:*

*נשים לב כי ההתבדרות איננה סימטרית: וכן כי תמיד מתקיים , ולפיכך יש שוויון בין האינפורמציה המתואמת של מקיים שהוא המרחק בין ה-joint probability האמיתי שלהם לבין ההסתברות שלהם בהנחה שהם בלתי תלויים:*

*העודף של אורך הודעה הנדרש להשתמש ב- יכול לעבור אופטימיזציה של הקוד ע"י שימוש ב-.*

*מאפיינים של אנטרופיה יחסית:*

*ומכאן נובע כי-*

*כמו כן מתקיים-*

1. *אי שוויון האינפורמציה- כאשר אם"ם .*
2. *חוסר השליליות של אינפורמציה מתואמת- כאשר השוויון מתקיים אם"ם בלתי תלויים. כאשר יש שוויון בהתפלגות בין* *ל- אשר .*
3. *התנייה מורידה את האנטרופיה- .*
4. *גבול חוסר התלות- .*
5. *אי שוויון ינסן- אם קמור אז .*

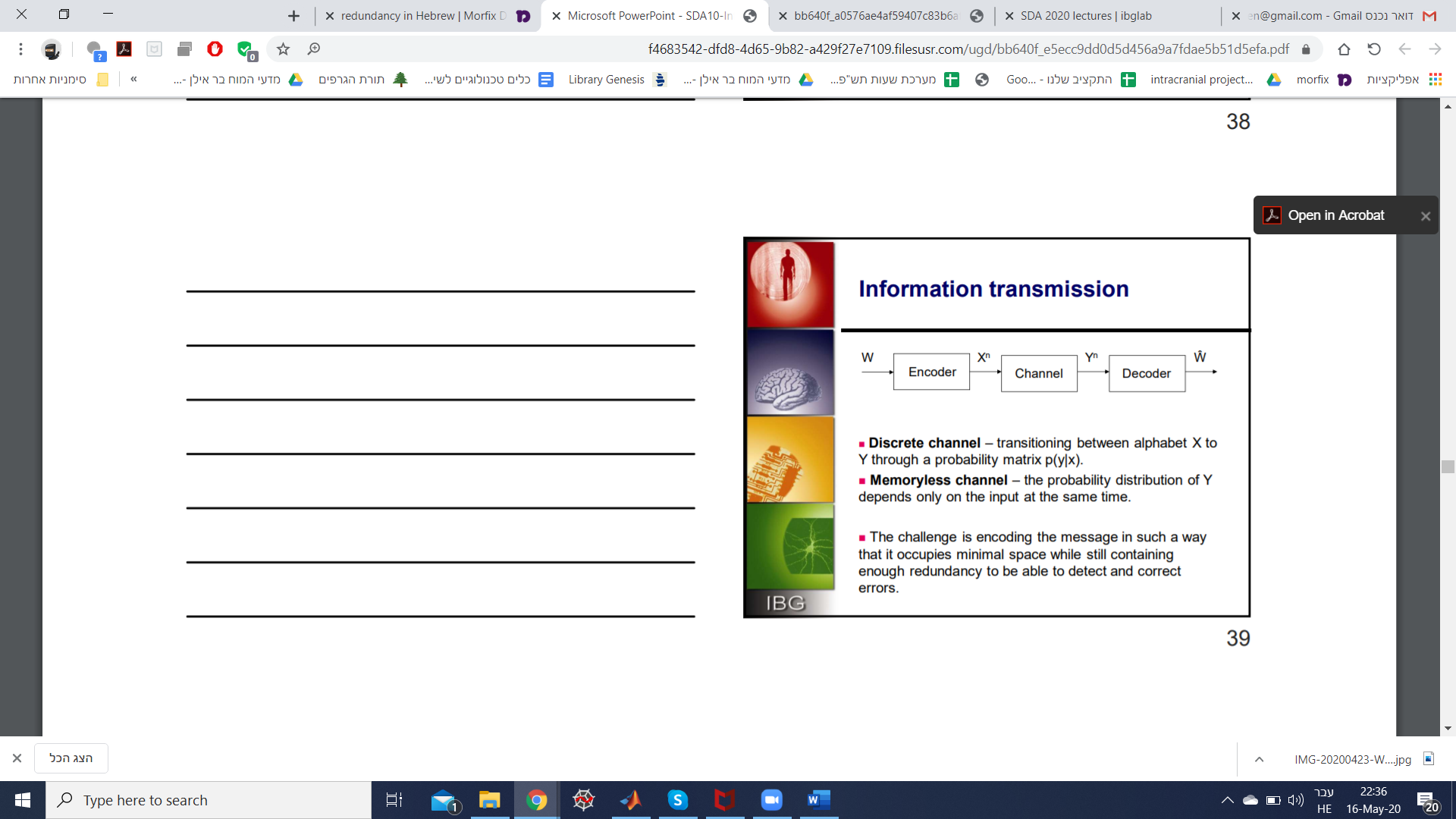
***תשדורת של אינפורמציה***

*הבעיה שאיתה התמודד שאנון כשפיתח את התורה הוא עסק בשאלות על מידת הרעש של המערכת וכיצד אפשר להתמודד איתה עם העברת אינפורמציה. השאלה מתכתבת באופן ישיר עם שאלות על מערכת העצבים- כיצד ניתן להתמודד עם כך שנוירונים רועשים ועדיין לאפשר העברת מידע.*

*דוגמא לכך היא למשל העברת תמונה שיש בה 1000 על 1000 פיקסלים, ואנחנו רוצים לשמור את התמונה אבל לא רוצים לשמור מיליון ביטים. בתהליך כיווץ/דחיסה אנחנו מבררים את האנטרופיה של משתנה (למשל תמונה) ותאפשר לנו לשמור את המידע עד חסם מסויים שהוא חסם האנטרופיה. כך למשל פועל מכשיר הפקס שמעביר מידע על כל נקודה בדף- האם היא שחורה או לבנה ומעבירים מידע על מספר ביטים ברצף שהם אפסים או אחדים ולא מעבירים ביט-ביט. הדור הראשון של הפקסים למשל העביר רק את מספר האפסים שיש למשל בכל שורה. בתהליך הדחיסה צריך להיות פרוטוקול זהה בין הפקסים שיבטיח שהפקסים "יבינו" אחד את השני. מחקרים שלמים עוסקים רק בשאלות איך שני צדדים יודעים להבין אחד את השני ובמערכת העצבים לשאלה הזאת יש ביטוי אפילו חד יותר- איך נוירונים יכולים להבין אחד את השני. בשלב הזה אנחנו לא נעבוד עם השאלות הללו, אלא רק נתעסק רק במה שקורה בתווך בין שני הצדדים בהינתן שהם יודעים לתקשר.*

*בכל מקרה נשים לב כי דחיסה (compression) הפוכה מתמסורת (transmission)- בעוד שהראשונה מצמצמת את כל העודף שיש ב-data, בתמסורת מוסיפים עודף ל-data על מנת לאפשר תיקוני טעויות. למשל באינטואיציה מהיומיום, כשאנחנו משתמשים בקו משובש, אנחנו מבקשים מאדם לחזור על הדברים שלו לפחות פעם אחת כדי לוודא מה אמר. כך למשל במערכת העצבים עולה השאלה כמה פעמים צריך לשדר שוב ושוב את אותו המסר כדי לוודא שהוא יגיע או להוסיף על גבי המסר מידע שיבטיח error correction.*

*ניתן לשרטט את הטרנספורמציות בין השלבים השונים של התמסורת באופן הבא. יש מילה אותה רוצים להעביר, צריך לקודד אותה (למשל למשהו בינארי) ל-, מעבירים אותה דרך ערוץ ומתקבל והצד השני צריך להשתמש בפענוח ולקבל חזרה את . כשהערוץ איננו ערוץ מושלם, יש פה בעיה הכוללת מגוון רחב של כלים מתמטיים לניתוח שלה. אנחנו נכיר את הבעיה הזו באופן מצומצם- רק תוך התייחסות לערוצים דיסקרטיים המאפשרים לנו לייצג את המעבר בין הערכים השונים באמצעות מטריצה.*



*בערוץ (channel) דיסקרטי עושים המרה בין אלפבית לאלפבית באמצעות מטריצת הסתברות . מה ההסתברות לקבל 0 בהינתן שנכנס 0 ומה ההסתברות לקבל 1 בהינתן שנכנס 0 וכו'. כמו כן נשתמש במונח נוסף- ערוץ חסר זיכרון, שבו התפלגות ההסתברות של תלויה אך ורק בקלט שהתקבל באותו הזמן. אין היסטוריה של מה שהיה קודם, ולכן פשוט יותר למדל זאת (אין צורך לזכור אם הביט הקודם השתבש או לא). האתגר הוא לקודד את המסר באופן כזה שדורש כמה שפחות מקום כאשר עדיין צריך לאפשר מספיק עודף על מנת לזהות ולתקן טעויות.*

*דוגמאות לערוצים-*

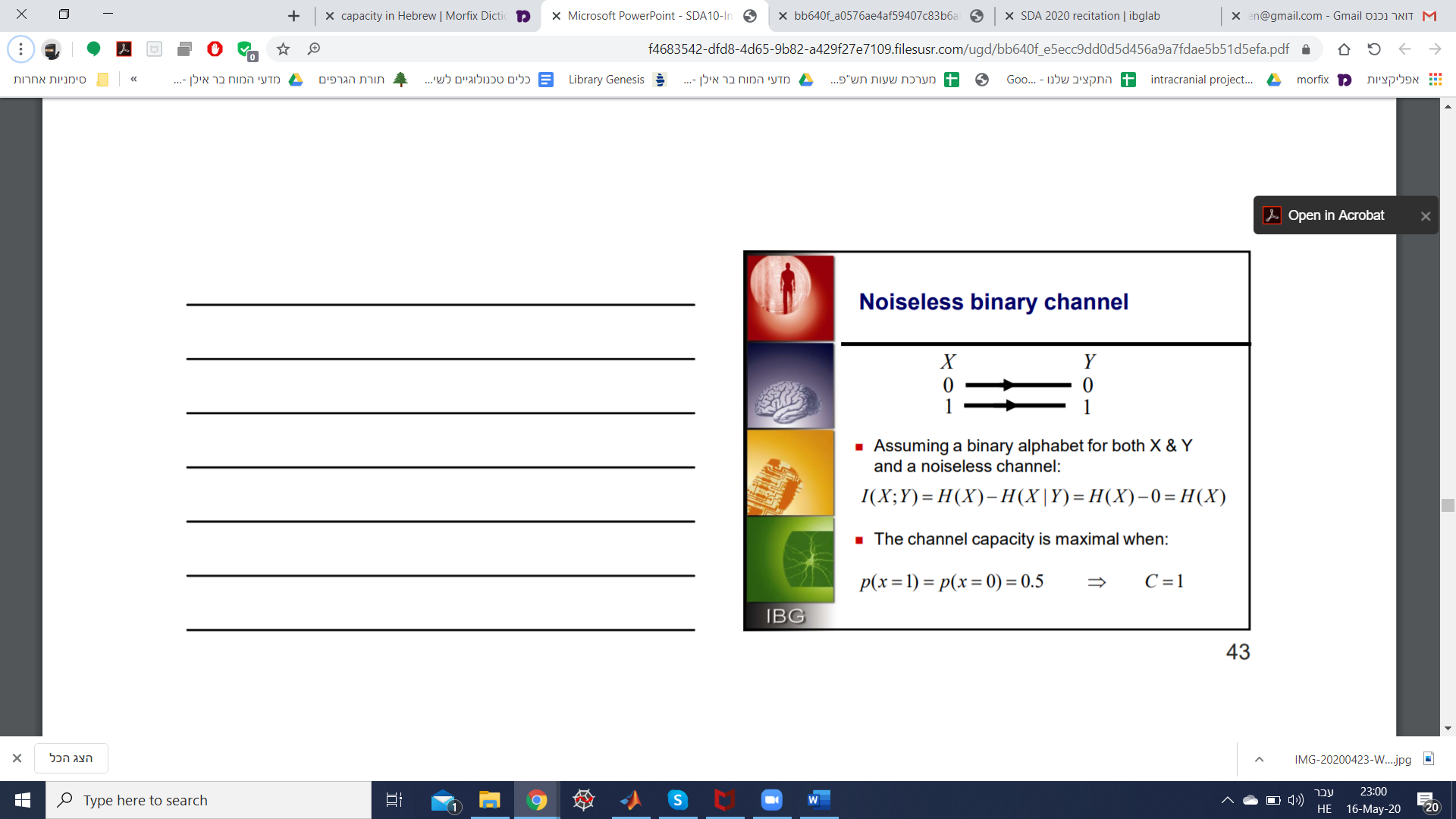
1. *מילה באנגלית יכולה להיות מתורגמת לרצף של הברות דרך דיבור אשר עוברות דרך האוויר (הערוץ שלנו) ועם השמיעה מתורגמת בחזרה לסדרה של הברות ומאפשרת ליצור מחדש את המילה. איכות הערוץ מושפעת ממגוון דברים- הרעש הסביבתי (למשל אם יושבים בפאב), האם יש קיר בין שני הדוברים, רוח ועוד.*
2. *מילה W באנגלית יכולה להיות מתורגמת לרצף של אותיות המיוצגות באמצעות 8 ביטים של קוד ASCII ועוברות דרך קו תמסורת וכאשר הן מתקבלות במחשב אחר הן עוברות תרגום בחזרה לסדרת אותיות ובונות את המילה. מה יכול לגרום לערוץ כזה להיות לא אמין? שדה מגנטי חזק, פגיעת ברק, הליכה על כאבלים ובגדול כל בעיה חשמלית.*

*לערוצים יש מספר מאפיינים:*

* *לכל ערוץ יש קצב שידור- מספר הסמלים שהערוץ יכול לשדר ליחידת זמן. למשל אם נסתכל על הדיבור שלנו נוכל להתבונן במספר ההברות שאנחנו יודעים לייצר בדקה- קצב השידור האישי של כל אחד. גם לרשתות תקשורת יש את הקצב הזה וכשמוכרים למשל אינטרנט דרך ספקית 500 מגה-ביט מתאים ל-500 ביטים שיכולים לעבור בשנייה אחת לראוטר. קצב השידור נקבע בהתאם למאפיין התקשורת- הברות ליחידת זמן למשל.*
* *לערוצים יש קצב השגיאה- אשר קובעים לכל סמל ספציפי את ההסתברות שסמל אחר יתורגם לאחר הערוץ. למשל אם אנחנו יושבים בפאב ומנסים לדבר ויש הרבה רעש סביבנו, קצב השגיאה יהיה גבוה וייצג את מספר המקרים שהברות התבלבלו ביניהן (לרוב מיוצג ע"י הסתברות).*
* *קצב השגיאה של הערוץ קובע את הקיבולת שלו- מספר הביטים של אינפורמציה שמועברים לכל סמל שנשלח. למשל אם כל סמל מיוצג ע"י ביט אחד אז אפשר להעביר 1 או 0. זו תהיה השאלה המרכזית שתהיה חשובה לנו- מה כמות הביטים שאפשר להעביר בהינתן ערוץ מסויים.*
* *כאשר מכפילים את קצב השידור והקיבולת זה בזו מקבלים את קצב המידע (data rate), הקצב שבו אינפורמציה יכולה להישלח בערוץ.*

*לחישוב של קיבולת הערוץ עבור ערוץ דיסקרטי חסר זיכרון, הקיבולת מגבילה את קצב העברת האינפורמציה שניתן להעביר עם הסתברות שגיאה קטנה מאוד. כלומר הקיבולת מוגבלת ע"י האינפורמציה המתואמת הממוקסמת ע"י התפלגות הקלט והפלט שלו (צמצום קצב השגיאה).*

*נפתח את המקרה הפשוט ביותר- ערוץ בינארי ללא רעשים נראה כך:*



*בהנחה של אלפבית בינארי גם עבור וגם עבור וערוץ ללא רעש, מתקיים:*

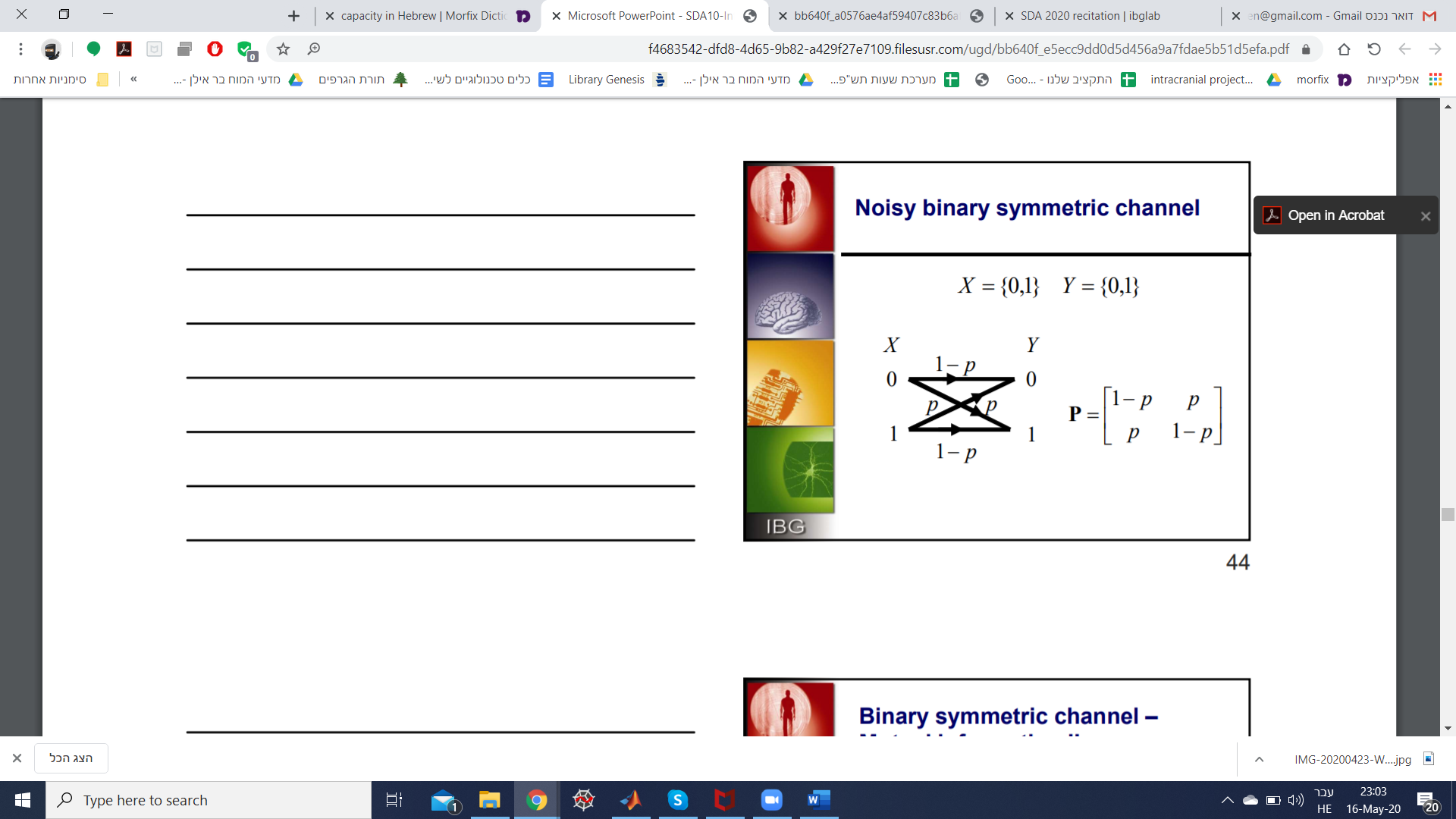
*אם אין רעש אז כאשר מגלים לנו את הפלט יודעים בדיוק מה היה הקלט ולכן אין כל חוסר ודאות ומתקיים ולכן:*

*במצב זה הקיבולת של הערוץ מקסימלית כאשר מקסימלית, וזה קורה כזכור מהדוגמא של אנטרופיה לשתי תוצאות של הטלת מטבע כאשר:*

*ומכאן:*

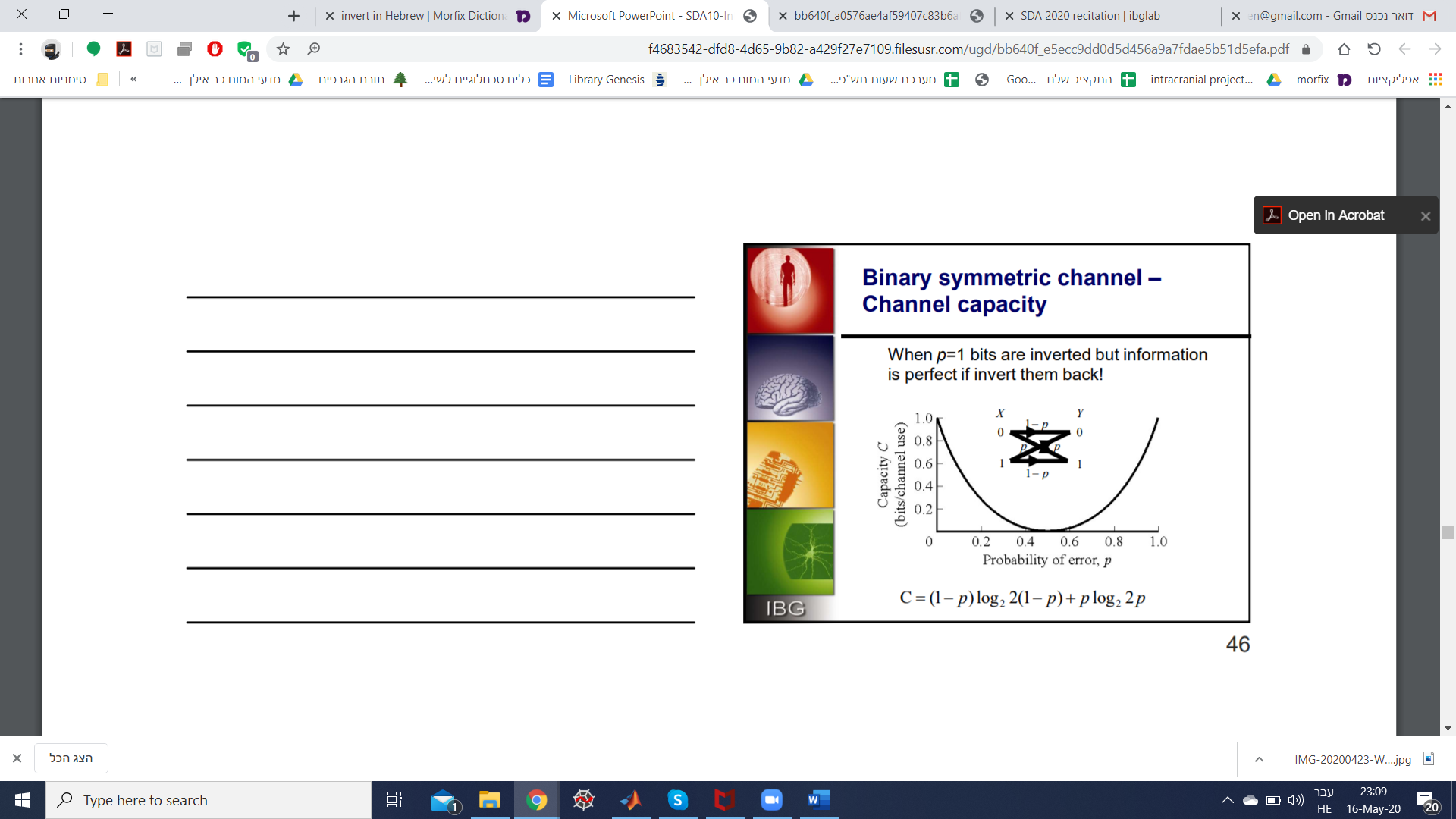
*מהווה חסם עליון ובמקרה שבו אפילו אם אין רעש ולמשל נכנס קלט רק של , מקבלים וגם .*

*ניתן להביע את הערוץ ע"י מטריצה- כאשר ולכן כדי להביע את הרעש בערוץ נקבל:*



*כמו כן בערוץ בינארי סימטרי מתקיים עבור אינפורמציה מתואמת:*

*מייצג את האנטרופיה של הסתברויות השגיאה בקלט שלנו. שוויון מתקיים כאשר ולכן במצב כזה מתקיים גם השוויון , ומכאן הקיבולת של הערוץ היא . נשים לב כי במצב שבו יש לנו טעות בעלת הסתברות לכל קלט, מתקבל כי ולכן , אולם עבור כל מספר אחר יכול לעבור מידע בערוץ- ייתכן שבאופן יותר איטי ופחות מדוייק מ-. כאשר אז ביטים עוברים טרנספורמציה, אך האינפורמציה מושלמת עם נעשה את הטרנספורמציה ההופכית.*



*אם נציב את הנוסחא לאנטרופיה, נקבל את הנוסחא הבאה עבור קיבולת:*

***דוגמא לחישוב קיבולת:***

*נתון כי הקלטים והפלטים הם בינאריים, וחוסר הוודאות המקסימלי של הקלט הינו . בהינתן שיש שגיאת מדידה בשיעור :*

*כלומר אם אנחנו משדרים 1000 ביטים של מידע אז 919 מתוכם יתקבלו. שאנון טען כי במצב כזה עדיין אפשר להביא למצב של העברת המידע המלא הזה. לדוגמא ע"י שליחת כל ביט פעמיים שתביא לירידה בשגיאה- מ-0.01 ל-0.0001. אולם כמובן במצב כזה קצב השידור יורד והקיבולת יורדת- ולכן שאנון טען שאפשר למצוא קידוד שהוא יותר יעיל. לפי שאנון לא ניתן לעלות הקיבולת של 0.9192 מכיוון שהיא האינפורמציה המקסימלית של הקלט/פלט. תאורטית אפשר להגיע ל-0.9192 בדיוק, וכיום ספרים רבים עוסקים בשיטות שהולכות ומשתפרות להעברת מידע (פרוטוקולים, תיקון רעשים, תיקון שיבושים וכדו').*

***איך מתמודדים עם שגיאות?***

*בהנחה שאנחנו יודעים שהולכות להיות שגיאות, כיצד אנחנו יכולים להיות בטוחים שהאינפורמציה הועברה? שאנון טען כי ניתן להקטין את השגיאה כרצוננו ע"י תיקונים- ייתכן במחיר הקיבולת, אך זה אפשרי. אנחנו נוכל להבטיח שביטים יעברו עד כדי הסתברות קטנה כרצוננו. אם אנחנו חסרי מזל אנחנו לא יכולים להיות בטוחים, אך אנחנו יכולים לוודא שאנחנו יכולים להתמודד עם אחוז מסויים של טעויות. למשל בזיהוי תקיפת סייבר מסוג מסויים (adversarial attack) אנחנו מחפשים את המקרים בהם ההסתברות לקבל מידע מסויים דרך ערוץ היא נמוכה מאוד. מהי אחת הדרכים עבורנו להיות מסוגלים שאנחנו יכולים לזהות כל שגיאה יחידה במסר שלנו? איך אנחנו יכולים לוודא שאנחנו מסוגלים לתקן שגיאה בהודעה?*

*אנחנו יכולים לפעול טוב יותר- להתקרב לקיבולת הערוץ ככל הניתן על אף שייתכן שבמקרה כזה נצטרך להשתמש באופן קבוע בהודעות ארוכות. קיבולת הערוץ מוגדרת בתור המידע שעובר דרך הערוץ, ולכן אם אנחנו צודקים בהגדרה שלנו לאינפורמציה, אנחנו יכולים לקבל מדידה מדוייקת לתשובה לשאלה כמה ביטים אנחנו יכולים לשלוח בערוץ שלנו. באופן אינטואיטיבי, קיבולת של ערוץ זה דבר הגיוני- מתחילים עם חוסר וודאות מקסימלי לגבי הסמל שנכנס לערוץ, וחוסר הוודאות יורד כשאנחנו רואים שהסמל יצא.*

***משפט הצפנת הערוץ***

*בהינתן קוד שהוא:*

* *בעל אינדקס מהקבוצה*
* *יש לו פונקציית קידוד*
* *ויש פונקציית פענוח*

*אז השיעור של הקוד הוא , וניתן להגיע לשיעור זה אם קיים רצף המובילה לשגיאה אינפיניטיסימלית כאשר . כל השיעורים מתחת לקיבולת של ערוץ הם ברי השגה ולכן .*

***משתנים רציפים***

*עד כה דיברנו על משתנים בדידים ע"י מדידת האינפורמציה באמצעות במצב שבו יש לנו מצבים. אולם, למספר ממשי יש אינסוף ביטים, ולכן באופן תיאורטי יש לו אינפורמציה אינסופית. למשתנים רציפים יש את היתרונות שלהם כיוון שהרבה פעולות מתמטיות אפשר לבצע רק במצב רציף (למשל נוירון שקצב הירי שלו מתפלג נורמלית). לכן אנחנו מעוניינים באיזשהו מדד שמאפשר לנו להתמודד עם ההבחנה בין שני נוירונים נורמליים בעלי שגיאות מדידה שונים.*

*הכלי שלנו להתמודד עם זאת הוא תחת ההנחה שתמיד יש רעש או קוונאטיזציה של המספר שמגדירה סדרת שלבים הנבדלים זה מזה. כל העברה למחשב כוללת בתוכה קוואנטיזציה וכל מכשיר מדידה שלנו כולל רעש מסויים שלא מאפשר לנו להבחין בין שני ערכים קרובים.*

*אולם, אם בכל זאת ננסה לדון באנטרופיה עבור משתנה רציף, נשתמש בצפיפות ההסתברות במקום בהסתברות ואז נקבל אנטרופיה המתארת זאת-*

*חשוב לשים לב כי עבור האנטרופיה מתפלגת במצב כזה ולכן משתמשים באנטרופיה הדיפרנציאלית (משתמשים בסימבול דומה להפתעה , אך הוא מוגדר באופן שונה). כשהוספנו את נפטרנו מההתבדרות וקיבלנו סכום על גבי צפיפות ההסתברות ולכן:*

*ניתן להגדיר אנטרופיה באופן שונה- אנטרופיה דיפרנציאלית:*

*דוגמא 1- התפלגות אחידה (במקטע ):*

*אם למשל אז האנטרופיה הדיפרנציאלית היא ואם אז היא , וזה מסתדר באופן הגיוני עם כך שככל שהטווח רחב יותר כך האנטרופיה הדיפרנציאלית יותר גדולה (יודעים פחות). אולם, בשונה מהגבלת האי-שליליות של האנטרופיה הקודמת, עבור האנטרופיה הדיפרנציאלית שלילית במילים אחרות האנטרופיה הדיפרנציאלית שלנו יכולה לקבל כל מספר בין .*

*דוגמא 2- התפלגות נורמלית ()*

*גם כאן ניתן לקבל אנטרופיה דיפרנציאלית שלילית כאשר .*

*אנחנו צריכים לנסות להבין מה זה אומר כאשר האנטרופיה היא שלילית, מה המשמעות מבחינת ביטים. נתבונן בדוגמא כיצד ניתן להתמודד עם מצב כזה ונחזור לביטוי כפי שהוגדר קודם עבור , האנטרופיה הכוללת את . למשל האנטרופיה של משתנה רציף שנדגם מתבססת על כך קוונטיזציה של ביטים של משתנה (כלומר דיוק של ) ונתונה ע"י:*

*לדוגמא בהתפלגות אחידה על פני המקטע עם רזולוציה של , עבורה מתקבל:*

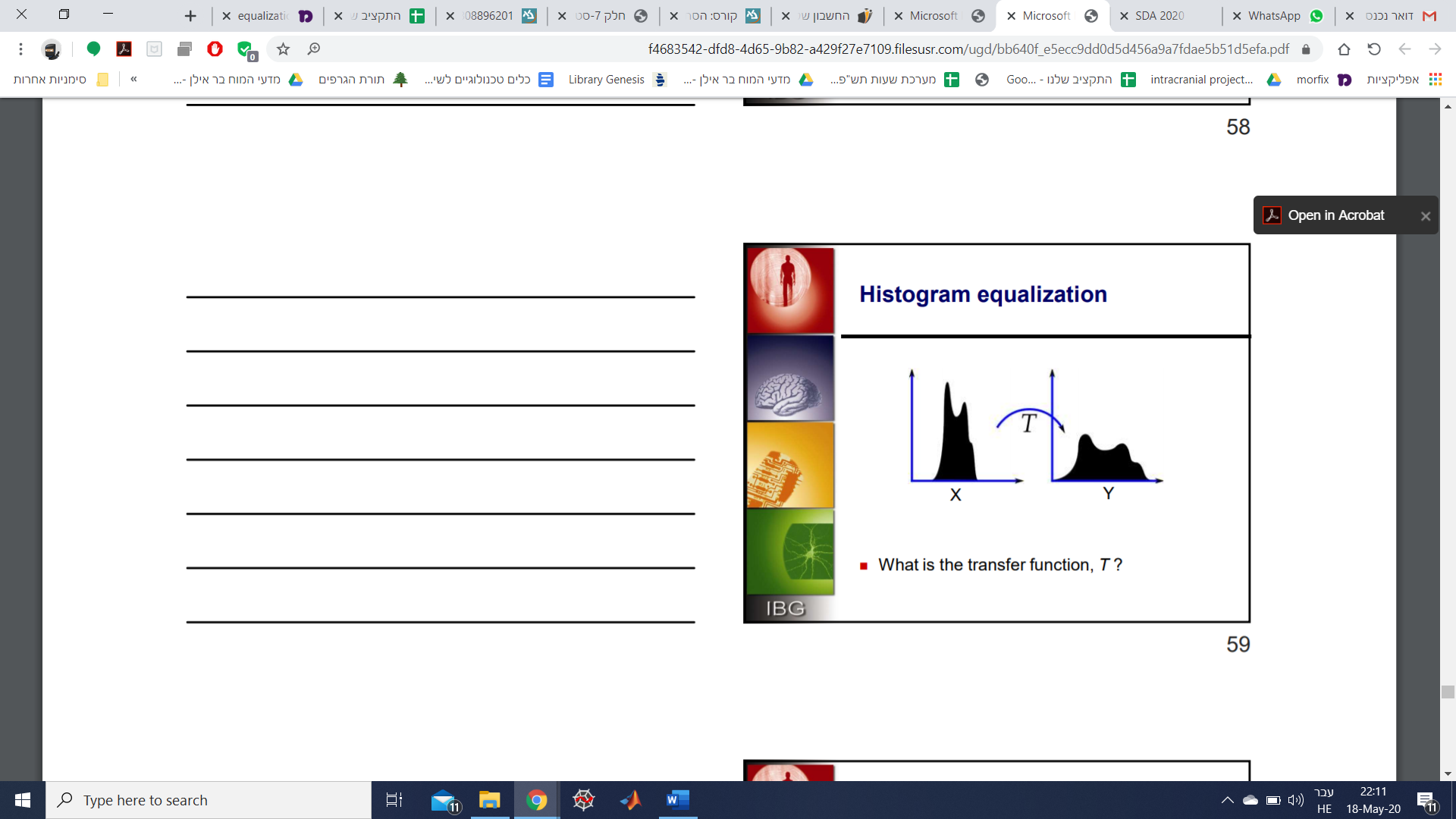
*כמו כן בהתפלגות אחידה על פני המקטע עם רזולוציה של מתקבל:*

*וזאת כיוון ששני הביטים הראשונים הם תמיד .*

***נוירונים ואנטרופיה***

*ניתן להעלות שאלות על בסיס תורת האינפורמציה בתחום הנוירופיזיולוגיה כמו למשל- כמה אינפורמציה נוירונים מעבירים, כמה אינפורמציה מועברת בספייק יחיד (בהסתכלות בינארית- היה ספייק או לא האינפורמציה היא 1, בהסתכלות על האם היה ספייק במסגרת של פרק זמן ספציפי- אנטרופיה אינסופית לרזולוציה אינסופית או ברזולוציה של זמן בדיד כתלות ברזולוציה- למשל למילישנייה האינפורמציה היא 10 לכל שנייה), כמה פעילות של ספייק נותנת לנו מידע על גירוי (כמה אנטרופיה נשארת בהתנהגות האורגניזם השלם אם מורידים נוירון אחד- כמו בדוגמא עם הקוף), האם הייצוג הנוירונלי אופטימלי, האם האינפורמציה המקודדת ע"י אוכלוסיית נוירונים כוללת עודפים (למשל האם מספיקים רק 10% מהנוירונים על מנת להביע את אותו המידע), האם קצב ירי מכיל את כל האינפורמציה ואם כן מה הגבול התאורטי של אינפורמציה במערכת העצבים. בעשורים שבין 1990 ל-2010 הנושאים הללו היו מרכזיים בתחומי המחקר במדעי המוח, בעיתונים דוגמת nature ו-science עסקו בדיוק בשאלות הללו.*

*נתבונן בכמה חישובים וכיצד ניגשים אליהם. אם קצב הירי הוא הקידוד במערכת ניתן לחשב את האנטרופיה המקסימלית, אז לכל קצבי הירי יש הסתברות שווה. לנוירון בעל קצב ירי מתקיים לכל קצב ירי: . למשל אם יש לנו טווח בין 0 ל-63, האנטרופיה המקסימלית שיש לנו תהיה 6 שהוא . כאשר הנוירון שלנו יתנהג לפני פונקציית דלתא, כלומר קצב ירי קבוע, האנטרופיה תהיה מינימלית. אולם, כאשר קצב הירי מקבל קלט שהוא משתנה לא אחיד נוסף, נקבל שילוב בין הערכים האמיתיים של הנוירון לבין הקלט שקיבל שיכול למשל להתמקד באיזור מסויים. אחד הכלים שמאפשרים לנו לעשות את זה נקרא histogram equalization המאפשרת לנו לשטח היסטוגרמה.*



*אנחנו נרצה לחפש מהי פונקציית הטרנספורמציה . אחד הכלים לעשות זאת הוא לנסות להעביר את הערכים לטווח יותר רחב ע"י החישוב:*

*אם לא מתפלג באופן אחיד, אז צריך להתאים לפונקציה שעולה מונוטונית :*

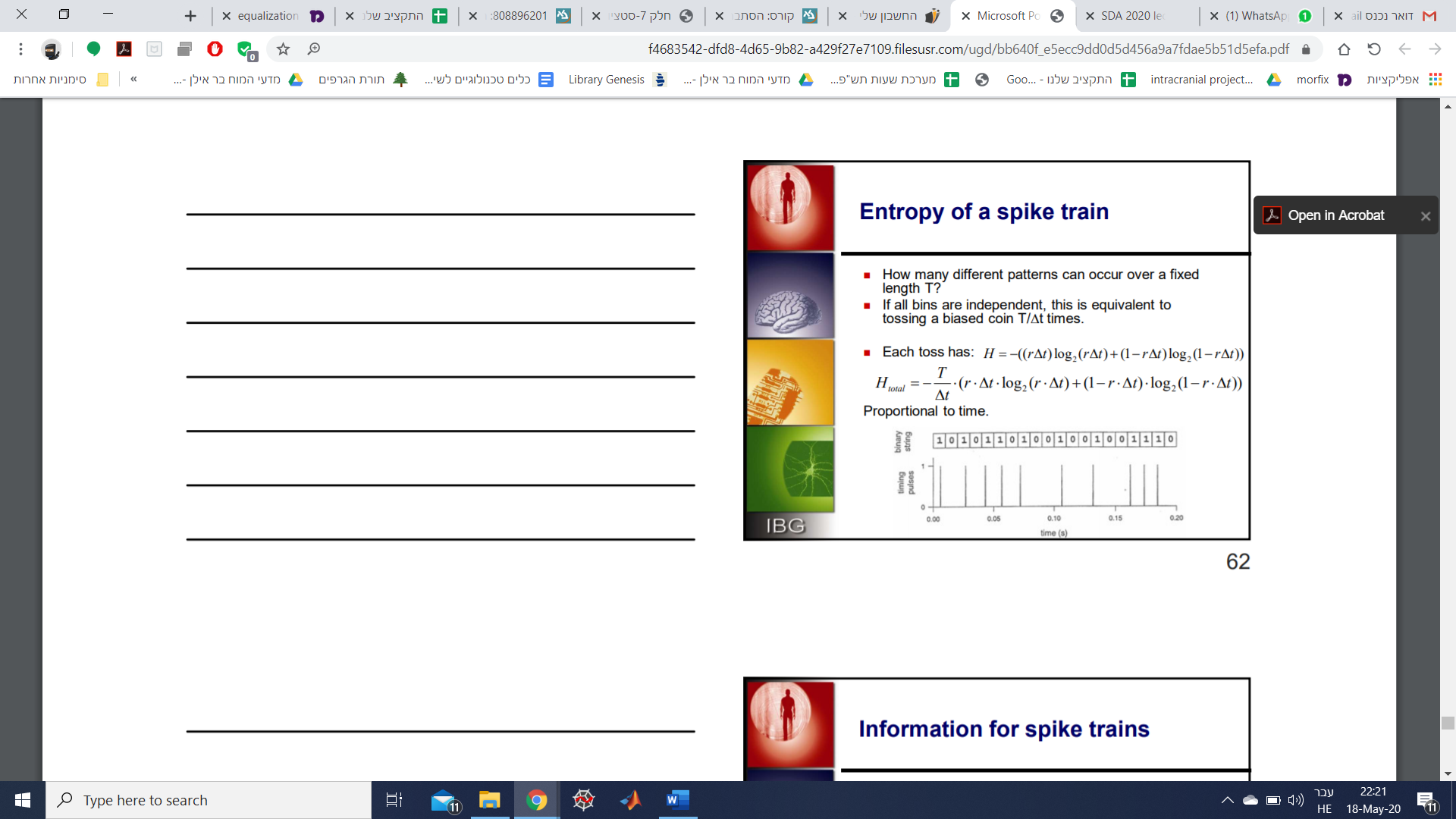
*ניתן להוסיף עוד ביטים לאיזורים עם הסתברות גבוהה יותר.*

*דוגמא נוספת עוסקת באנטרופיה המקסימלית של קידוד אוכלוסייה. ננסה לחשב את האנטרופיה המקסימלית ע"י מציאת האנטרופיה המקסימלית עבור כל נוירון באוכלוסיה זו. אם למשל שני נוירונים יורים באותו קצב ירי ממוצע הם זהים לנוירון בודד שיורה פי שתיים יותר זמן, מה שמוביל לחישוב אנטרופיה שנמצאת ביחס ישיר למספר הנוירונים:*

*החסם התחתון, אם הם זהים, אז האנטרופיה מייצגת רק אחד מהם . החסם העליון קורה אם"ם ו- בלתי תלויים. סוג כזה של קידוד בלתי תלוי לרוב נקרא "factorial code". כיום יש מחקרים בתחום שעוסקים בשאלה איזה מידע ניתן לקודד ע"י קורלציה בין נוירונים (דוגמת המחקרים של משה אבלס, שמראים שבאיזורים מסויימים במוח המידע עובר רק ע"י הקורלציה ולא ע"י הספייקים).*

*חישוב נוסף שניתן לבצע הוא לאנטרופיה של סדרת ספייקים (spike train) ובבסיסו עומדת השאלה כמה תבניות שונות יכולות לקרות עבור אורך קבוע. אם כל הבינים הם בלתי תלויים, השאלה הזו זהה לשאלה על הטלת מטבע בעל הטייה פעמים. בכל הטלה מתקיים:*

*כלומר פרופורציונלי לזמן .*



*חישוב האינפורמציה בסדרת ספייקים מצריכה להתחשב בכל תבנית של ספייקים על פני מקטע בתור מספר יחיד בינארי. ישנם הרבה מספרים בינאריים אפשריים- יכול להיות מורכב להעריך את אלא אם קצר במיוחד. אינפורמציה של קצב הירי היא מספר הביטים לשנייה (או ביטים לספייק) כפי שהם קשורים לקלט. אם ההסתברות לספייק בבין היא קטנה (קצב ירי נמוך או קצב דגימה גבוה) אנחנו יכולים להעריך את האנטרופיה של קצב הירי בתור:*

*כמו כן ניתן להתבונן בשאלה על קידוד- הזמן של ספייק מול ספירת מספר הספייקים לשנייה. שאלה זו משמעה מהי האינפורמציה המקסימלית שמועברת תוך שימוש במדידת spike count לעומת . לדוגמא- בהנחה שלנוירון יש תקופה רפרקטורית בת 3 מילישניות, מהי האנטרופיה המקסימלית בהינתן 10 בינים ברצף בני 3 מילישניות כל אחד שבכל אחד מהם יכול להיות ספייק אחד לבין מול בין אחד בן 30 מילישניות שיכול להכיל עד מידע על עד 10 ספייקים? שאלה זו מובילה לשאלה על הנחת היסוד ממנה התעלמנו עד כה- האם נוירון מעביר מידע כאשר הוא איננו יורה.*