

## מבוא לביולוגיה מערכתית - תרגיל מספר 4

### שאלה 1: שערך פרמטרים

הגן X מתבטא בתאי חיידקים בקצב ייצור של 2 molecules/min, וקצב פירוק של 0.04 1/min. בתגובה לשינוי בהרכב הסוכרים במדיום הגידול של התאים (זמן 0), עלה קצב הייצור שלו ל- 8 molecules/min.

א. (10 נקודות) כתבו סימולציה של רמות הביטוי של הגן X לאורך 3 שעות מזמן השינוי (זמן 0), עבור דגימה בכל 30 דקות, כאשר המדידה היא: (1) 10 חזרות על המדידה מאוכלוסיית תאים: רעש gaussian עם שונות 10, (2) 10 מדידות של תאים בודדים: רעש poisson עם שונות 2. שרטטו את תוצאות הסימולציה בגרף.

ב. (10 נקודות) השתמשו בנתונים שייצרתם בסימולציה על מנת לשערך את קצב הייצור וקצב הפירוק של הגן. שערכו את ערכי הפרמטרים: (1) בנפרד בכל אחת מהמדידות הבודדות של הניסוי (2) באמצעות שימוש בנתונים מתוך 3 מדידות בכל פעם. שרטטו בגרף את ערכי השערך של כל פרמטר, והשוו אותם לערך האמיתי באמצעותו ייצרתם את הנתונים.

ג. (10 נקודות) דונו בקצרה בהתאמה בין ערכי הפרמטרים המשוערים לבין הערכים האמיתיים, על פי כל אחד מהמודלים לרעש.

ד. (10 נקודות) במקום לאסוף 3 דגימות בכל נקודת זמן, הגדילו את תדירות הדגימה בניסוי פי 3 (כל 10 דקות), ודגמו דגימה בודדת. כמות הדגימות הכוללת לא השתנתה. בשערך הפרמטרים, האם יש עדיפות לדגום 3 פעמים את אותה נקודת זמן, או לדגום ללא חזרות בתדירות פי 3 יותר גבוהה?

**טיפ:** קיימות פונקציות מובנות במטלב לדגימה מהתפלגויות שונות. randn() להתפלגות נורמלית, poissrnd() לפואסונית.

העזרו בקוד לדוגמה הנמצא בחלק של parameter estimation במודל. (למשתמשי מטלב: אם פונקציית האופטימיזציה בדוגמה לא עובדת בצורה טובה נסו פונקציות אופטימיזציה שונות שאפשר למצוא בdocumentation של מטלב [fmincon, lsqcurvefit, nlinfit]).

### שאלה 2: שערך ממדידות

בקובץ rna\_levels.xlsx נתונות מדידות של רמות הביטוי של 10 גנים, שנאספו מתאים של מערכת החיסון בתגובה לסטימולציה עם גורם חיידקי. המדידות נלקחו עם 3 חזרות.

א. (5 נקודות) שערכו את קצב הייצור וקצב הפירוק של כל אחד מהגנים (עבור כל גן, לא עבור כל חזרה) מתוך המדידות הנתונות באמצעות המודל הבסיסי של ייצור ופירוק רנ"א שנלמד בכיתה.

ב. (10 נקודות) השוו את מידת ההתאמה של המדידות של כל אחד מהגנים הנתונים למודל שבניתם ולפרמטרים שהערכתם. אילו גנים מתאימים יותר למודל, ואילו פחות?

ג. (5 נקודות) מה יכולות להיות הסיבות להבדל בהתאמת הגנים למודל?

ד. (10 נקודות) נסוך להגיע להתאמה טובה יותר למדידות על ידי אחת ההרחבות של המודל הבסיסי שנלמדו בכיתה. תארו את ההרחבות שהוספתם למודל ועבור אילו גנים הן משפרות את ההתאמה.

**טיפ:** העמודות באקסל מייצגות מדידה בכל חצי שעה לאורך 5 שעות. לכל גן קיימות 3 שורות שמייצגות 3 חזרות שונות. השתמשו בפונקציה xlsread כדי לקרוא את הדאטא לתוך מטלב (או פונקציה מקבילה בפייתון). העזרו במצגת התרגול הראשונה בשביל הסינקטס למניפולציה של המטריצה.

### שאלה 3: אנליזה של רצפים

החלבונים FOXH1, SMAD2 הם פקטורי שעתוק הפעילים במהלך ההתפתחות העוברית של בעלי חוליות. הפקטור SMAD2 מפעיל ביטוי גנים, ואילו הפקטור FOXH1 מדכא ביטוי של גנים.

א. (10 נקודות) בניסוי קישור in-vitro זוהה קישור של שני הפקטורים הנ"ל לסט הרצפים המופיעים בקבצים FOXH1.fa ו-SMAD2.fa (בהתאמה לשמות הפקטורים). השתמשו בנתונים אלה על מנת לבנות מודל PSSM של אתר הקישור ב-DNA של פקטורי השעתוק הנ"ל. צרו שני logo עבור המוטיבים: אחד המתאר את שכיחות הנוקלאוטידים, ושני המתאר את האינפורמציה במוטיבים.

ב. (10 נקודות) בקובץ fish\_promoters.fa מופיעים רצפי פרומוטרים של גנים של דגים. השתמשו במודלים שבניתם על מנת לאתר את גני המטרה של כל אחד מהפקטורים הנ"ל. הגדירו סף של  $pvalue < 1e-3$ . דרגו אותם לפי חוזק הקישור.

ג. (10 נקודות) דונו בקצרה אילו מוטיבים נוצרים ברשת השעתוק ששחזרתם באמצעות האנליזה הנ"ל.

**טיפ:** הוראות שימוש בכלי alignment תוכלו למצוא בpdf המצורף.

Meme מוצא כברירת מחדל 3 מוטיבים. זה שמוטיב כלשהו הוא לא המוטיב הכי שכיח ברצפי המטרה לא בהכרח אומר שהוא לא מעניין.