

מעבדה לסטטיסטיקה 2024, מטלה 6

הצגה ב 25.6

אמידת מספר עותקים מדגימה אחת

אנחנו נניח שהנתונים מיוצרים על פי המודל הבא:

$$Y_i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = a_i \cdot f(gc_i) \cdot \eta_i$$

כאשר:

- Y_i הכיסוי הנצפה בתא ה*i*.
- $a_i \in \{0.5, 1, 1.5, 2\}$ מספר העותקים הגנומי (copy number) בתא ה*i*.
- $f(gc_i)$ היא פונקציית התוחלת התלויה ב*gc* של התא
- $\eta_i \sim N(1, v)$ שגיאה כפלית לא מוסברת (בקירוב 1) עם v קטן.

1. עבור המודל הזה, כתבו פונקציה שאומדת מספר עותקים גנומי. הפונקציה תקבל כקלט וקטור Y , וקטור GC, ופונקציה f (תוצאה של רגרסיית splint), ותוציא וקטור אומדים ל*a*. (כדאי לוודא שהפונקציה מסתדרת עם פלט באורך 1).

2. בחרו אזור באורך 1000 תאים רצופים. השתמשו בפונקציית f שאמדתם במטלה 5 כדי לאמוד את מספר העותקים בתאים אלו.

a. הציגו את ההתפלגות השולית ואת ההתפלגות לאורך הגנום של האומדים שהתקבלו.

b. מצאו מדד מספרי שמודד על כמה מספרי העותקים שנאמדו קרובים ל1.

3. העריכו את איכות האומד באופן מילולי. האם יש בו בעיות פוטנציאליות?

4. (רשות) הציעו רעיון כיצד להעריך את הפרמטר v (השונות של η_i).