תרגיל 1

אלה פאליק 208882134

1 במאי 2020

שאלה 1: Poisson loading

מסתכלים על תהליך הכנסת תאים לטיפות. נניח שיש לנו שני סוגי תאים - בני אדם ועכברים, במספר שווה. מתכננים להעביר אותם ביחס $\lambda = \text{rate of } \frac{\#cells}{\#droplets}$. אוספים M טיפות ומרצפים אותם. קיבלנו

$$\begin{aligned} N_H &= \text{\#droplets with human cells only} \\ N_M &= \text{\#droplets with mouse cells only} \\ N_B &= \text{\#droplets with both human and mouse cells} \end{aligned}$$

סעיף א

נניח $M = 10,000$ ו $\lambda = 0.1$. מה התוחלות של N_H, N_M, N_B?

נתון שיש מס' שווה של תאי בני אדם ועכברים, לכן ההסתברות לתא אנושי = $\alpha = \frac{1}{2}$.

נסמן $C = \text{\#cells in droplet}$. נזכור שראינו בהרצאה $C \sim Po(\lambda)$. אז לכל $k \geq 0$, $\mathbb{P}(C = k) = \frac{e^{-\lambda}}{k!}\lambda^k$.

נסמן $X_{H,i}$ מציין של האם בטיפה ה i יש רק תאים אנושיים. אז $N_H = \sum_{i=1}^{M} X_{H,i}$.

נסתכל על טיפה כלשהי i. הסתברות לתא אנושי = $\alpha = \frac{1}{2}$ אז $\mathbb{P}(X_{H,i} = 1|C = k) = \left(\frac{1}{2}\right)^k$ לכל $k > 0$. אם אין תאים בטיפות לא נחשיב את התא, $\mathbb{P}(X_{H,i} = 1|C = 0) = 0$. לכן נקבל מנוסחת ההסתברות השלמה:

$$\begin{aligned} \mathbb{P}(X_{H,i} = 1) &= \sum_{k=0}^{\infty} \mathbb{P}(X_{H,i} = 1|C = k) \cdot \mathbb{P}(C = k) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \cdot \frac{e^{-\lambda}}{k!}\lambda^k = \\ &= \sum_{k=1}^{\infty} \left(\frac{\lambda}{2}\right)^k \cdot \frac{e^{-\lambda}}{k!} + \left(\frac{\lambda}{2}\right)^0 \cdot \frac{e^{-\lambda}}{0!} - e^{-\lambda} = \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \cdot \frac{e^{-\lambda}}{k!} - e^{-\lambda} = \\ &= e^{-\frac{\lambda}{2}} \sum_{k=0}^{\infty} \left(\frac{\lambda}{2}\right)^k \cdot \frac{e^{-\frac{\lambda}{2}}}{k!} - e^{-\lambda} = e^{-\frac{\lambda}{2}} \underbrace{\sum_{k=0}^{\infty} \mathbb{P}\left(Po\left(\frac{\lambda}{2}\right) = k\right)}_{=1} - e^{-\lambda} = e^{-\frac{\lambda}{2}} - e^{-\lambda} \end{aligned}$$

סה"כ נקבל מלינאריות התוחלת:

$$\mathbb{E}[N_H] = \mathbb{E}\left[\sum_{i=1}^{M} X_{H,i}\right] = \sum_{i=1}^{M} \mathbb{E}[X_{H,i}] = \sum_{i=1}^{M} \mathbb{P}(X_{H,i} = 1) = M \cdot \left(e^{-\frac{\lambda}{2}} - e^{-\lambda}\right)$$

באופן דומה, מכיוון שההסתברות לתא עכבר = $1 - \alpha = \frac{1}{2}$, נקבל $\mathbb{E}[N_M] = M \cdot \left(e^{-\frac{\lambda}{2}} - e^{-\lambda}\right)$.

נסמן $N = \#\text{non empty droplets}$. אז $N_H + N_M + N_B = N$ לכן

$$\mathbb{E}\left[N_B\right] = \mathbb{E}\left[N\right] - \mathbb{E}\left[N_H\right] - \mathbb{E}\left[N_M\right]$$

נשים לב $N = M - \#\text{empty droplets}$. נחשב:

$$\begin{aligned}\mathbb{E}\left[\#\text{empty droplets}\right] &= \sum_{i=1}^{M}\mathbb{E}\left[\text{droplet i is empty}\right] = \sum_{i=1}^{M}\mathbb{P}\left(\text{droplet i is empty}\right) = \\ &= \sum_{i=1}^{M}\mathbb{P}\left(C=0\right) = M\cdot e^{-\lambda}\end{aligned}$$

לכן $\mathbb{E}\left[N\right] = M - \mathbb{E}\left[\#\text{empty droplets}\right] = M - Me^{-\lambda}$. אז

$$\mathbb{E}\left[N_B\right] = \mathbb{E}\left[N\right]-\mathbb{E}\left[N_H\right]-\mathbb{E}\left[N_M\right] = M-Me^{-\lambda}-2M\cdot\left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right) = M\left(1-2e^{-\frac{\lambda}{2}}+e^{-\lambda}\right) = M\left(1-e^{-\frac{\lambda}{2}}\right)^2$$

נציב את המספרים הנתונים $M = 10,000$ ו $\lambda = 0.1$ ונקבל:

$$\boxed{\mathbb{E}\left[N_H\right] = \mathbb{E}\left[N_M\right] = 463.92, \qquad \mathbb{E}\left[N_B\right] = 23.78}$$

סעיף ב

נניח כעת $M \sim Po\left(10,000\right)$. איך זה משנה את התוצאות?
הבעיה היא שלא נוכל להשתמש בלינאריות התוחלת ישירות, אז נשתמש קודם בנוסחת התוחלת השלמה:

$$\mathbb{E}\left[N_H\right] = \sum_{m=0}^{\infty}\mathbb{E}\left[N_H|M=m\right]\cdot\mathbb{P}\left(M=m\right)$$

$M \sim Po\left(10,000\right)$. נסמן $\delta = 10,000$. אז אנחנו יודעים $\mathbb{P}\left(M=m\right) = \frac{e^{-\delta}}{m!}\delta^m$. עבור כל m מסעיף א, נקבל $\mathbb{E}\left[N_H|M=m\right] = m\cdot\left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right)$. לכן סה"כ

$$\begin{aligned}\mathbb{E}\left[N_H\right] &= \sum_{m=0}^{\infty}\mathbb{E}\left[N_H|M=m\right]\cdot\mathbb{P}\left(M=m\right) = \sum_{m=0}^{\infty}m\cdot\left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right)\cdot\frac{e^{-\delta}}{m!}\delta^m = \left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right)\delta\sum_{m=1}^{\infty}\frac{e^{-\delta}}{(m-1)!}\delta^{m-1} = \\ &= \left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right)\delta\sum_{m=0}^{\infty}\frac{e^{-\delta}}{m!}\delta^m = \left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right)\delta\underbrace{\sum_{m=0}^{\infty}\mathbb{P}\left(Po\left(\delta\right)=m\right)}_{1} = \left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right)\delta\end{aligned}$$

באותו אופן נקבל שוב $\mathbb{E}\left[N_M\right] = \mathbb{E}\left[N_H\right] = \left(e^{-\frac{\lambda}{2}}-e^{-\lambda}\right)\delta$.
כעת $N = M - \#\text{empty droplets}$. שוב בעזרת נוסחת התוחלת השלמה:

$$\mathbb{E}\left[\#\text{empty droplets}\right] = \sum_{m=0}^{\infty}\mathbb{E}\left[\#\text{empty droplets}|M=m\right]\cdot\mathbb{P}\left(M=m\right)$$

מסעיף א, $\mathbb{E}\left[\#\text{empty droplets}|M=m\right] = me^{-\lambda}$. לכן סה"כ (חישוב דומה לזה שלמעלה):

$$\mathbb{E}\left[\#\text{empty droplets}\right] = \sum_{m=0}^{\infty}\mathbb{E}\left[\#\text{empty droplets}|M=m\right]\cdot\mathbb{P}\left(M=m\right) = \sum_{m=0}^{\infty}me^{-\lambda}\cdot\frac{e^{-\delta}}{m!}\delta^m = \delta e^{-\lambda}$$

לכן $\mathbb{E}[N] = \mathbb{E}[M] - \mathbb{E}[\#\text{empty droplets}] = \delta - \delta e^{-\lambda}$. אלו אותן תוצאות כמו סעיף א, רק שהחלפנו את M ב δ לכן נקבל $\mathbb{E}[N_B] = \delta\left(1 - e^{-\frac{\lambda}{2}}\right)^2$.
עכשיו $\delta = 10,000$ וזה שווה ל M בסעיף א', וקיבלנו את אותן תוצאות בהחלפת M ב δ, לכן **אין שינוי** בתוצאות.

סעיף ג

נניח כעת שאנחנו לא יודעים את λ. האם אפשר לשערך אותו מתצפיות? כתבו את הפונקצית L והוציאו גרף של LL מול ערכי λ שונים.
האם אפשר למצוא פתרון אנליטי?
אילו ערך λ הכי סביר אם נתון $M = 10,000, \quad N_H = 629, \quad N_M = 678, \quad N_B = 45$?

$$L(\lambda | N_H, N_M, N_B) = \mathbb{P}(N_H, N_M, N_B | \lambda)$$

הרכב התאים בטיפות שונות הוא ב"ת. נסמן

$$\begin{aligned} X_{H,i} &= 1_{\{\text{droplet i with human cells only}\}} \\ X_{M,i} &= 1_{\{\text{droplet i with mouse cells only}\}} \\ X_{H,i} &= 1_{\{\text{droplet i with both human and mouse cells}\}} \end{aligned}$$

בהינתן λ, ראינו כי $\mathbb{P}(X_{H,i} = 1) = \mathbb{P}(X_{M,i} = 1) = e^{-\frac{\lambda}{2}} - e^{-\lambda}$. בנוסף ראינו $\mathbb{P}(\text{empty cell}) = e^{-\lambda}$. אז

$$\begin{aligned} \mathbb{P}(X_{B,i} = 1) &= 1 - \mathbb{P}(X_{H,i} = 1) - \mathbb{P}(X_{M,i} = 1) - \mathbb{P}(\text{empty cell}) = \\ &= 1 - 2e^{-\frac{\lambda}{2}} + 2e^{-\lambda} - e^{-\lambda} = 1 - 2e^{-\frac{\lambda}{2}} + e^{-\lambda} = \left(1 - e^{-\frac{\lambda}{2}}\right)^2 \end{aligned}$$

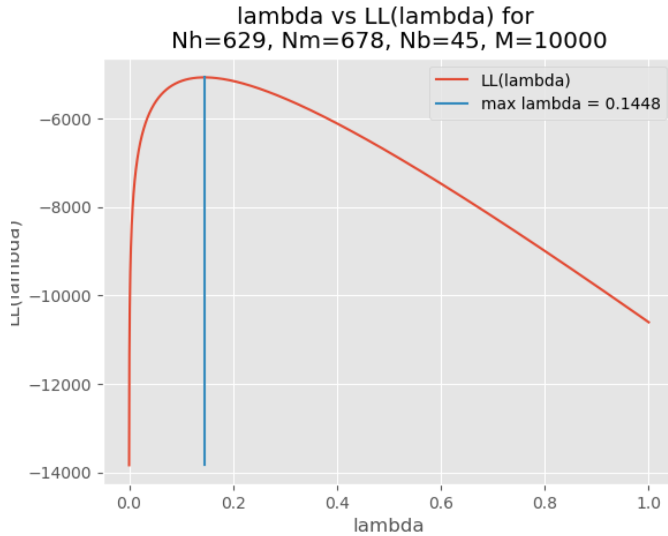
נשים לב כי ההסתברות לא תלוי בסדר הטיפות. אז סה"כ נקבל:

$$\begin{aligned} L(\lambda | N_H, N_M, N_B) &= \mathbb{P}(N_H, N_M, N_B | \lambda) = \\ &= \mathbb{P}(X_{H,i} = 1)^{N_H} \mathbb{P}(X_{M,i} = 1)^{N_M} \mathbb{P}(X_{B,i} = 1)^{N_B} \mathbb{P}(\text{empty cell})^{M - N_H - N_M - N_B} = \\ &= \left(e^{-\frac{\lambda}{2}} - e^{-\lambda}\right)^{N_H} \left(e^{-\frac{\lambda}{2}} - e^{-\lambda}\right)^{N_M} \left(\left(1 - e^{-\frac{\lambda}{2}}\right)^2\right)^{N_B} \left(e^{-\lambda}\right)^{M - N_H - N_M - N_B} = \\ &= \left(e^{-\frac{\lambda}{2}} - e^{-\lambda}\right)^{N_H + N_M} \left(1 - e^{-\frac{\lambda}{2}}\right)^{2N_B} \left(e^{-\lambda}\right)^{M - N_H - N_M - N_B} = \\ &= \left(e^{\lambda}\left(e^{-\frac{\lambda}{2}} - e^{-\lambda}\right)\right)^{N_H + N_M} \left(e^{\frac{\lambda}{2}}\left(1 - e^{-\frac{\lambda}{2}}\right)\right)^{2N_B} \left(e^{-\lambda}\right)^M = \\ &= \left(e^{\frac{\lambda}{2}} - 1\right)^{N_H + N_M} \left(e^{\frac{\lambda}{2}} - 1\right)^{2N_B} \left(e^{-\lambda}\right)^M = \boxed{\left(e^{\frac{\lambda}{2}} - 1\right)^{N_H + N_M + 2N_B} \left(e^{-\lambda}\right)^M} \end{aligned}$$

נעשה לוג:

$$LL(\lambda | N_H, N_M, N_B) = \log\left(\left(e^{\frac{\lambda}{2}} - 1\right)^{N_H + N_M + 2N_B} \left(e^{-\lambda}\right)^M\right) = \boxed{(N_H + N_M + 2N_B)\log\left(e^{\frac{\lambda}{2}} - 1\right) - \lambda M}$$

נתון לנו $M = 10,000, \quad N_H = 629, \quad N_M = 678, \quad N_B = 45$.
נוציא פלוט של ערכי λ מול $LL(\lambda | N_H, N_M, N_B)$:



איור 1: פלוט של λ מול $LL(\lambda)$ שחישבנו עבור $N_B = 45, \quad N_M = 678, \quad N_H = 629, \quad M = 10,000$.

כאן נקבל מקסימום ב $\lambda = 0.1448$.
נחשב פתרון אנליטי: נגזור את $LL(\lambda|N_H, N_M, N_B)$ ונשווה לאפס:

$$LL(\lambda|N_H, N_M, N_B)' = (N_H + N_M + 2N_B)\frac{1}{e^{\frac{\lambda}{2}} - 1} \cdot e^{\frac{\lambda}{2}} \cdot \frac{1}{2} - M = 0$$

$$\iff 1 + \frac{1}{e^{\frac{\lambda}{2}} - 1} = \frac{e^{\frac{\lambda}{2}}}{e^{\frac{\lambda}{2}} - 1} = \frac{2M}{N_H + N_M + 2N_B}$$

$$\iff \frac{1}{e^{\frac{\lambda}{2}} - 1} = \frac{2M}{N_H + N_M + 2N_B} - 1 = \frac{2M - N_H - N_M - 2N_B}{N_H + N_M + 2N_B}$$

$$\iff e^{\frac{\lambda}{2}} = \frac{N_H + N_M + 2N_B}{2M - N_H - N_M - 2N_B} + 1$$

$$\iff \boxed{\lambda = 2\log\left(\frac{N_H + N_M + 2N_B}{2M - N_H - N_M - 2N_B} + 1\right)}$$

נציב את הנתונים ונקבל $\boxed{\lambda = 0.1448}$ בדיוק כמו שרצינו.

סעיף ד

נניח כעת $M \sim Po(10,000)$. איך זה משנה את התוצאות?

$$L(\lambda|N_H, N_M, N_B) = \mathbb{P}(N_H, N_M, N_B|\lambda)$$

ההסתברויות עבור תאים יחידים הן זהות. ההבדל היחיד הוא כמות התאים הריקים, שהיא $M - N_H - N_M - N_B$ ועכשיו זה מ"מ. נעשה כמו בסעיף ב' עם נוסחת ההסתברות השלמה על כל האפשרויות של M. אז

$$L(\lambda|N_H, N_M, N_B) = \mathbb{P}(N_H, N_M, N_B|\lambda) = \sum_{m=0}^{\infty}\mathbb{P}(N_H, N_M, N_B|\lambda, M = m)$$

מסעיף ג' לכל m, $\mathbb{P}\left(N_H,N_M,N_B|\lambda,M=m\right)=\left(e^{\frac{\lambda}{2}}-1\right)^{N_H+N_M+2N_B}\left(e^{-\lambda}\right)^m$ ו $M\sim Po\left(10,000\right)$ לכן $\mathbb{P}\left(M=m\right)=\frac{e^{-\delta}}{m!}\delta^m$ עבור $\delta=10000$. לכן סה"כ:

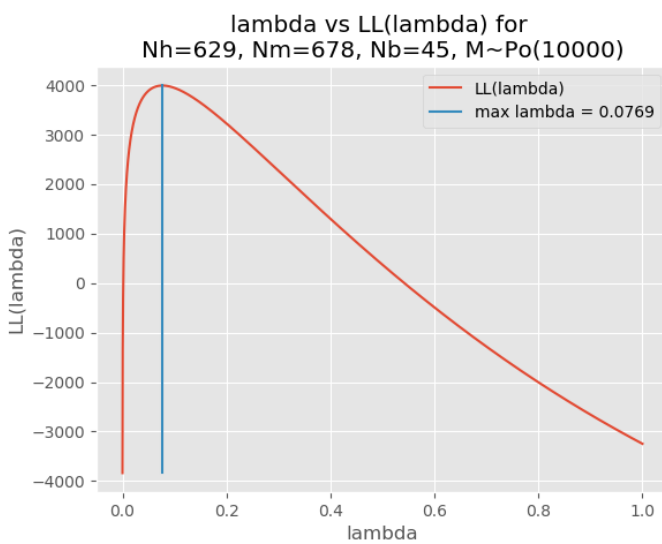
$$
\begin{aligned}
L\left(\lambda|N_H,N_M,N_B\right)=\mathbb{P}\left(N_H,N_M,N_B|\lambda\right)&=\sum_{m=0}^{\infty}\mathbb{P}\left(N_H,N_M,N_B|\lambda,M=m\right)\cdot\mathbb{P}\left(M=m\right)=\\
&=\sum_{m=0}^{\infty}\left(e^{\frac{\lambda}{2}}-1\right)^{N_H+N_M+2N_B}\left(e^{-\lambda}\right)^m\frac{e^{-\delta}}{m!}\delta^m=\\
&=\left(e^{\frac{\lambda}{2}}-1\right)^{N_H+N_M+2N_B}e^{-\delta+2\delta e^{-\lambda}}\overbrace{\sum_{m=0}^{\infty}\frac{e^{-\delta e^{-\lambda}}}{m!}\left(\delta e^{-\lambda}\right)^m}^{=1}=\\
&=\left(e^{\frac{\lambda}{2}}-1\right)^{N_H+N_M+2N_B}e^{-\delta}e^{2\delta e^{-\lambda}}
\end{aligned}
$$

נעשה לוג:

$$LL\left(\lambda|N_H,N_M,N_B\right)=\log\left(\left(e^{\frac{\lambda}{2}}-1\right)^{N_H+N_M+2N_B}e^{-\delta}e^{2\delta e^{-\lambda}}\right)=\boxed{\left(N_H+N_M+2N_B\right)\log\left(e^{\frac{\lambda}{2}}-1\right)-\delta+2\delta e^{-\lambda}}$$

נתון לנו $N_B=45\quad N_M=678,\quad N_H=629,\quad M=10,000,$.

נוציא פלוט של ערכי λ מול $LL\left(\lambda|N_H,N_M,N_B\right)$:



איור 2: פלוט של λ מול $LL\left(\lambda\right)$ שחישבנו עבור $N_B=45\quad N_M=678,\quad N_H=629,\quad M\sim Po\left(10000\right),$.

כאן נקבל מקסימום ב $\lambda=0.0769$.

ננסה לחשב פתרון אנליטי: נגזור את $LL\left(\lambda|N_H,N_M,N_B\right)$ ונשווה לאפס:

$$\begin{aligned}
LL\left(\lambda|N_H,N_M,N_B\right)' &= \left(N_H+N_M+2N_B\right)\frac{1}{e^{\frac{\lambda}{2}}-1}\cdot e^{\frac{\lambda}{2}}\cdot\frac{1}{2}-2\delta e^{-\lambda}=0\\
\iff \frac{e^{\frac{\lambda}{2}}}{e^{\frac{\lambda}{2}}-1} &= \frac{4\delta e^{-\lambda}}{N_H+N_M+2N_B}\\
\iff \frac{e^{\frac{3\lambda}{2}}}{e^{\frac{\lambda}{2}}-1} &= \frac{4\delta}{N_H+N_M+2N_B}
\end{aligned}$$

ולזה אין פתרון אנליטי (ממה שאני יודעת).

שאלה 2: Stochastic models of expression

מתואר המודל של שני מצבים עבור כמות רנ"א בתא.
נסתכל על שלוש אפשרויות:

1. מעברים מהירים: $\lambda = 1, \delta = 0.2, \kappa_{on} = 1, \kappa_{off} = 1$.
2. מעברים בינוניים: $\lambda = 1, \delta = 0.2, \kappa_{on} = 0.1, \kappa_{off} = 0.1$.
3. מעברים איטיים: $\lambda = 1, \delta = 0.2, \kappa_{on} = 0.01, \kappa_{off} = 0.01$.

ונשתמש באלגוריתם Gillespie כדי לסמלץ ריצה במודל.
יש לנו שתי אפשרויות לאסוף מדגם של מצבים:
- CollectSample - לקחת מצב אחרון מכל ריצה עבור n ריצות.
- CollectSamplesMultiRun: לקחת n מצבים מכל ריצה במרווחי זמן קבועים, עבור k ריצות.

סעיף א, ג

לכל אחד מאפשרויות המעברים: נדגום כמה ריצות ונצייר אותן. האם ההבדלים בהתנהגויות תואמים את האינטואיציה שלנו/מה שראינו בכיתה? הראו את ההתפלגות של R במצבי ON ו OFF, וההתפלגות השולית. האם זה תואם את מה שראינו בכיתה?
ראינו בהרצאה:
במצב שבו יש רק מצב on, $R \sim Po\left(\frac{\lambda}{\delta}\right)$.
עבור ההתפלגות הסטציונארית, כלומר כאשר t מספיק גדול $\mathbb{P}\left(on\right) = \frac{\kappa_{on}}{\kappa_{off}+\kappa_{on}}$.
$\mathbb{E}\left[dewll\, time\, on\right] = \frac{1}{\kappa_{off}}, \mathbb{E}\left[dewll\, time\, off\right] = \frac{1}{\kappa_{on}}$ תוחלת הזמן להישארות במצב.

מעברים מהירים

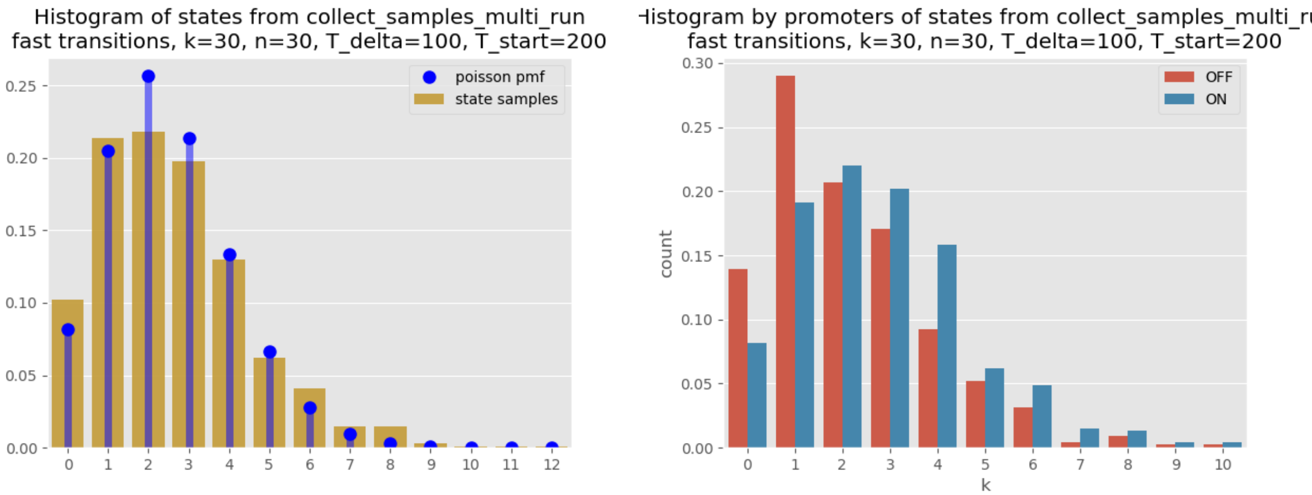
במקרה הזה החלפה מתרחשת מהר יותר משעתוק - מחכים $\frac{1}{\lambda}$ לשעתוק, אבל נשארים במצב on בממוצע $\frac{1}{\kappa_{off}}$ זמן. ראשית ב off אי אפשר לשעתק, אבל גם במצב on יהיו לנו מעט מאוד שעתוקים, כי נעדיף להחליף. לכן נצפה לראות R נמוך. מצד שני, מכיוון שאנחנו מחליפים כל הזמן לא יקרה מצב שבו נתקע במצב off, ולכן לא יהיה מצב שבו אנחנו רק מפרקים רנ"א בלי לאפשר שעתוק.

סימולציה של ריצה ניתן לראות עליות וירידות בלי העדפה ספציפית ל $k = 0$, וכן ערכי k קטנים יחסית בממוצע.

איור 3 : שינוי k בזמן סימולציה של ריצה אחת עם fast transitions עם $T_{end} = 10000$.

התפלגות R בשיעור דיברנו על זה שנצפה לקבל $\lambda_{effective} \approx \mathbb{P}(on) \cdot \lambda$, כלומר קצב השעתוק שנראה יהיה הקצב אם נסתכל רק על מה שקורה ב on. במקרה הזה המודל יראה בקירוב כמו המודל אם המכונה תמיד דולקת (כי אנחנו לא נתקעים ב off). ראינו בשיעור שעבור מודל רק עם מצב on נקבל $R \sim Po\left(\frac{\lambda}{\delta}\right)$. נכניס את קצב השעתוק שלנו ונקבל בקירוב $R \sim Po\left(\frac{\mathbb{P}(on) \cdot \lambda}{\delta}\right)$, כאשר $\mathbb{P}(on) = \frac{\kappa_{on}}{\kappa_{off}+\kappa_{on}} = \frac{1}{2}$.
ואכן, אם נדגום מתוך סימולציות על המודל, נוכל לראות שהקירוב הזה לא רע.

התפלגויות R **במצבי** on/off כאשר נסתכל על התפלגות R לפי סימולציות במצבי on ו off בנפרד, נוכל לראות את ההשפעה של מעברים מהירים : כמעט אין זמן לפעולות שאינן החלפה. ההתפלגות במצב off דומה למצב on כי אין מספיק זמן לפרק, וערכי k אליהם מגיעים באמת קטנים ביחס למעברים איטיים יותר, כי אין מספיק זמן לשעתק.



איור 4: היסטוגרמה של מצבים מסימולציות על המודל ע"י collect_samples_multi_run, עם fast transitions ופרמטרים $k = 30, n = 30, T_{\Delta} = 100, T_{start} = 200$, מול pmf של $Po\left(\frac{\mathbb{P}(on)\cdot\lambda}{\delta}\right)$ (בשמאל) ובחלוקה ל on/off (בימין).

מעברים איטיים

לעומת זאת, עבור מעברים איטיים, החלפה היא אירוע נדיר. שעתוק ופירוק מתרחשים משמעותית יותר מהחלפה, לכן בתוך כל אירוע של on נצפה לראות הרבה אירועי שעתוק. מצד שני כאשר אנחנו במצב off נוכל לחכות שם הרבה זמן. מכיוון שגם קצב הפירוק גדול מקצב ההחלפות וב off ניתן רק לפרק, אז במצב off נראה דעיכה, לעיתים עד ל 0, לפני החלפת מצב. כלומר, המודל מתנהג שונה לאורך זמן כתלות באיזה מצב אנחנו. בנוסף, מכיוון שאנחנו נשארים זמן רב באותו מצב אנחנו לא "מבזבזים" זמן על החלפות, ולכן נגיע די מהר להתפלגות סטציונארית.

סימולציה של ריצה ניתן לראות העדפה ברורה להישארות ב $k = 0$ (כאשר אנחנו תקועים ב off), וערכי k גדולים יחסית (כאשר אנחנו תקועים ב on).

איור 5: שינוי k בזמן סימולציה של ריצה אחת עם slow transitions עם $T_{end}=10000$.

התפלגות R ראינו בשיעור שבהתפלגות הסטציונארית נוכל לתאר את R בעזרת mixture distribution:

$$\mathbb{P}\left(R=k\right)=\mathbb{P}\left(on\right)\mathbb{P}^{on}\left(R=k\right)+\left(1-\mathbb{P}\left(on\right)\right)\mathbb{P}^{off}\left(R=k\right)$$

כאשר:

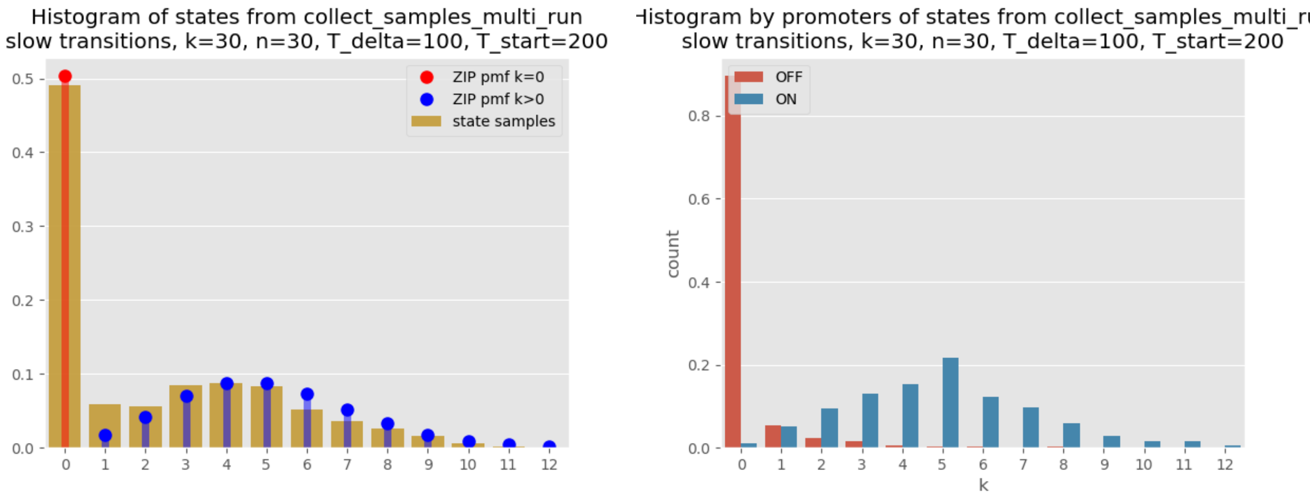
- $\mathbb{P}^{off}\left(R=k\right)$ הוא 1 עבור $k=0$ ו 0 אחרת, כלומר בהסתברות 1, במצב off נתייצב על 0 מולקולות, מכיוון שרק נפרק ולא נוכל לשעתק, ואין מספיק החלפות לכן נשאר ב off זמן רב, מספיק כדי להגיע ל 0.
- $\mathbb{P}^{on}\left(R=k\right)$ הוא, כמו שראינו, ה pmf של $Po\left(\frac{\lambda}{\delta}\right)$.
- $\mathbb{P}\left(on\right)=\frac{\kappa_{on}}{\kappa_{off}+\kappa_{on}}=\frac{1}{2}$.

לכן סה"כ נקבל קירוב של R בצורת zero inflated poisson (ZIP):

$$\mathbb{P}\left(R=k\right)=\begin{cases}k=0 & \left(1-\mathbb{P}\left(on\right)\right)+\mathbb{P}\left(on\right)\cdot e^{-\frac{\lambda}{\delta}}\\ k>0 & \mathbb{P}\left(on\right)\cdot\frac{1}{k!}\left(\frac{\lambda}{\delta}\right)^{k}e^{-\frac{\lambda}{\delta}}\end{cases}$$

ואכן אם נדגום מתוך סימולציות על המודל, שוב נקבל שהקירוב הזה לא רע.

התפלגויות R **במצבי** on/off כאן נוכל לראות בצורה משמעותית את העובדה שהמערכת נשארת במצב אחד זמן ממושך, ולכן מתנהגת כמו שתי מערכות שונות: במצב off נהיה בעיקר ב 0 - כי ההסתברות למעבר מאוד נמוכה, לכן תהיה נמוכה כמעט תמיד מההסתברות להוריד את k (כל עוד אפשר) שהיא $k\cdot\delta$. לכן במצב on נעדיף כמעט תמיד להוריד מאשר לעבור, ולכן כמעט לא נעבור ל off כל עוד k לא 0. לעומת זאת במצב on נראה התפלגות קרובה למה שהיינו רואים במערכת רק עם מצב on. ההתפלגויות מופרדות לגמרי.

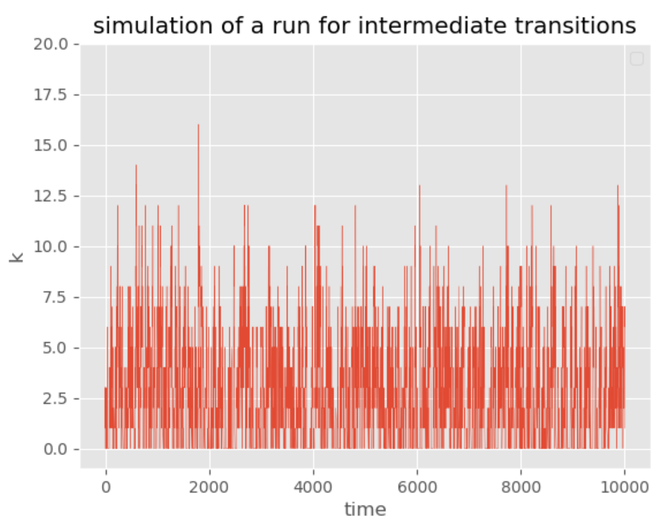


איור 6: היסטוגרמה של מצבים מסימולציות על המודל ע"י collect_samples_multi_run, עם slow transitions ופרמטרים $k = 30, n = 30, T_\Delta = 100, T_{start} = 200$, מול pmf של zero inflated poisson המתואר למעלה (משמאל), ובחלוקה ל on/off (בימין).

מעברים בקצב בינוני

לבסוף עבור מעברים בקצב בינוני, נרצה לקבל משהו "באמצע". עדיין יש יותר שעתוק ופירוק מאשר מעברים, אבל אנחנו עוברים מספיק כדי שלא נוכל להניח שבכל פעם שאנחנו ב off אנחנו מגיעים ל 0, ושכל פעם שאנחנו ב on אנחנו נמצאים שם מספיק זמן כדי לסמלץ מכונה שהיא רק on.

סימולציה של ריצה רואים כאן משהו ממוצע, כלומר יש הישארות ב $k = 0$ מידי פעם אבל בהחלט לא כמו במעברים איטיים, מצד שני מגיעים לערכי k גדולים יותר מאשר במעברים מהירים.



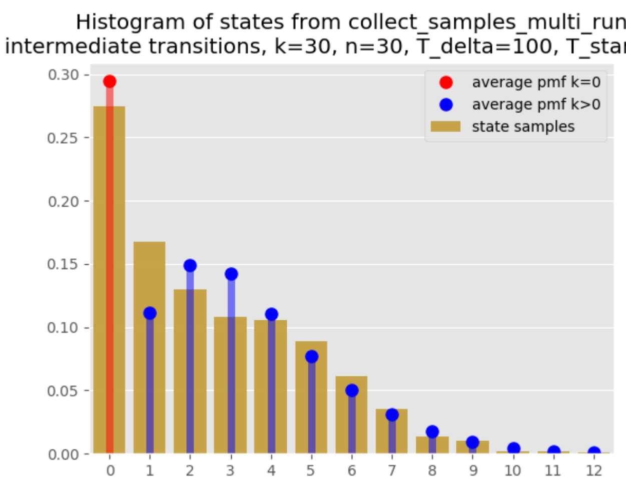
איור 7: שינוי k בזמן סימולציה של ריצה אחת עם intermediate transitions עם $T_{end} = 10000$.

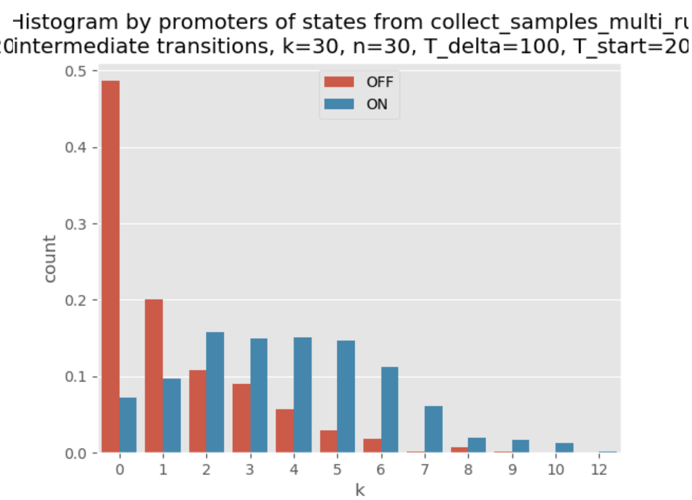
התפלגות R הרעיון שהגעתי אליו היה למצע בין שתי ההתפלגויות, כלומר

$$\mathbb{P}\left(R=k\right)=\begin{cases} k=0 & \frac{1}{2}\left(\left(1-\mathbb{P}\left(on\right)\right)+\mathbb{P}\left(on\right)\cdot e^{-\frac{\lambda}{\delta}}+e^{-\frac{\mathbb{P}(on)\cdot\lambda}{\delta}}\right)\\ k>0 & \frac{1}{2}\left(\mathbb{P}\left(on\right)\cdot\frac{1}{k!}\left(\frac{\lambda}{\delta}\right)^{k}e^{-\frac{\lambda}{\delta}}+\frac{1}{k!}\left(\frac{\mathbb{P}(on)\cdot\lambda}{\delta}\right)^{k}e^{-\frac{\mathbb{P}(on)\cdot\lambda}{\delta}}\right)\end{cases}$$

ובאמת נראה שזה קירוב לא רע, למרות שברור שהוא עובד פחות טוב מאשר למקרים הקיצוניים.

התפלגויות R **במצבי** on/off כאן אפשר לראות באופן יותר בולט (מאשר בהתפלגות השולית) את העובדה שזהו אכן מצב ביניים: במצב on נראה משהו דומה לפואסון שהיינו רואים במערכת שתמיד דולקת, אבל עם נטייה קלה לערכי k קטנים, ובמצב off לעומת זאת המערכת תנסה לרדת ל 0. העובדה שבמצב off נראה ערכי k יותר גבוהים כאן מאשר במעברים איטיים קשורה לכך שכאן יתקבלו יותר מעברים ל off בערכי k גבוהים מאשר במעברים איטיים שם נעבור ל off כמעט רק כאשר $k=0$.





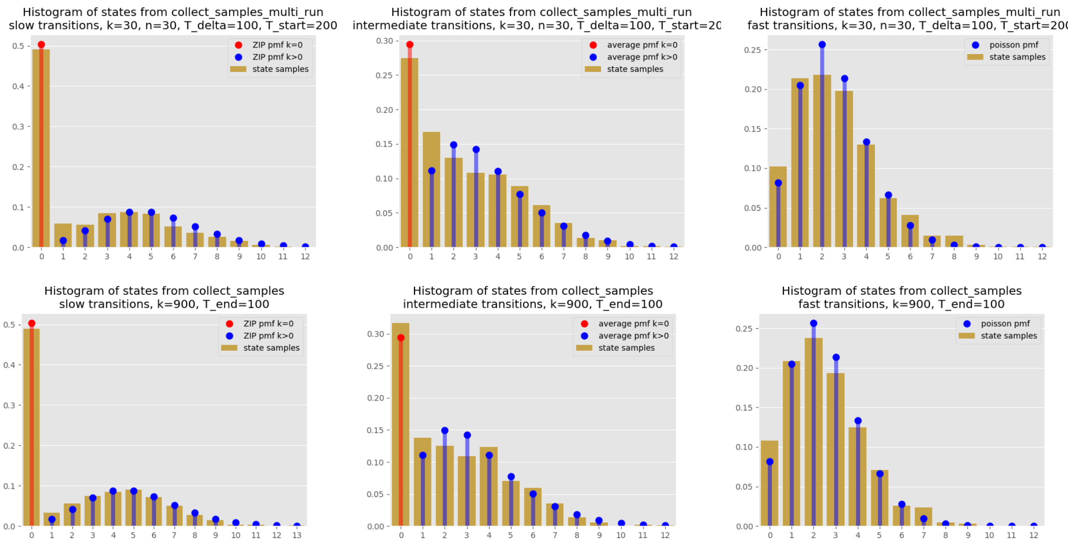
איור 8: היסטוגרמה של מצבים מסימולציות על המודל ע"י collect_samples_multi_run, עם intermidiate transitions ופרמטרים $k=30, n=30, T_{\Delta}=100, T_{start}=200$, מול pmf המתואר למעלה (בשמאל) ובחלוקה למצבי on/off (מימין).

סעיף ב

נעשה סימולציה של המערכת ונאסוף דוגמאות בעזרת שתי האופציות. נשים לב לטריידאוף בין לקיחת דגימות מרובות ממקטעים לעומת הרבה ריצות (n מול k).

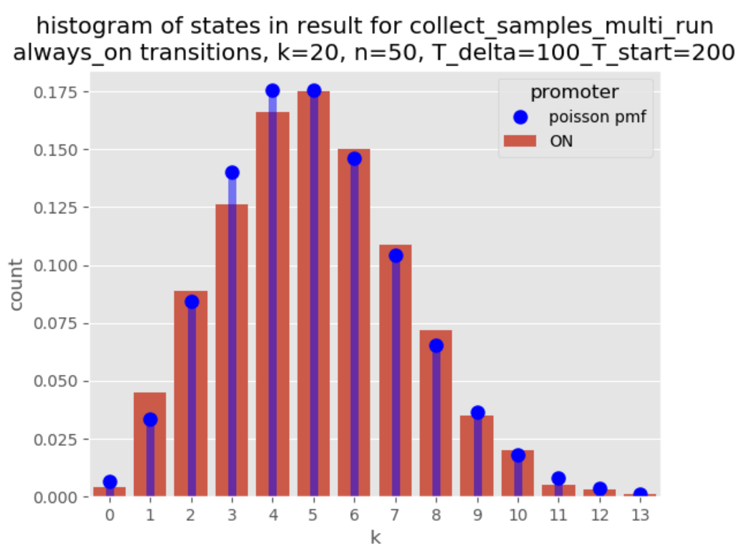
נזכור כי אם המערכת במצב ON תמיד, אז ההתפלגות של מספר הדוגמאות צריכה להתאים לתוצאות שראינו בכיתה.

כאשר אוספים דוגמאות בעזרת שתי האופציות לא ראיתי שינוי משמעותי. למשל:



איור 9: היסטוגרמה של מצבים מסימולציות על המודל ע"י collect_samples_multi_run (למעלה) וע"י collect_samples עבור פרמטרים $k=30, n=30, T_\Delta=100, T_{start}=200$ ל multirun ו $k=900, T_{end}=100$ לדגימות הרגילות (מה שיתן לנו אותו מספר דוגמאות).

בנוסף אם נותנים למערכת פרמטרים שישאירו אותה ב on תמיד (מתחילים מ on ונותנים קצב מעבר 0), אז אנחנו רוצים לקבל משהו שתואם מודל שבו אין לנו מצב off, וראינו שבמקרה כזה $R \sim Po\left(\frac{\lambda}{\delta}\right)$. ואכן מקבלים:



איור 10: היסטוגרמה של מצבים מסימולציות על המודל ע"י collect_samples_multi_run, עם always_on transitions ופרמטרים $k=20, n=50, T_\Delta=100, T_{start}=200$, מול pmf של $Po\left(\frac{\lambda}{\delta}\right)$.