Física Biológica

Probabilidades y paseo al azar.

1. (*tomado del libro "Molecular Driving Forces"*) Suponga que presentó una solicitud de ingreso a tres lugares, A, B y C. Suponga que las probabilidades de ser aceptadx son $p(A) = 0.10$, $p(B) = 0.30$ y $p(C) = 0.50$. Piense qué significan estas probabilidades en términos de un diagrama de conjuntos. Suponga que los eventos de aceptación son independientes.
   - (a) ¿Cuál es la probabilidad de que entre en algún lugar (es decir, que sea aceptadx en al menos un lugar)?
   - (b) ¿Cuál es la probabilidad de que B y C lx acepten?
2. (*tomado del libro "Molecular Driving Forces"*) Suponga que las cuatro bases A, C, T y G ocurren con la misma probabilidad en una secuencia de ADN de nueve monómeros.
   - (a) ¿Cuál es la probabilidad de encontrar la secuencia AAATCGAGT por azar?
   - (b) ¿Cuál es la probabilidad de encontrar la secuencia AAAAAAAAA mediante el azar?
   - (c) ¿Cuál es la probabilidad de encontrar una secuencia que tenga cuatro A, dos T, dos G y una C, como la de (a)?
3. (*tomado del libro "Molecular Driving Forces"*) Suponga un péptido del que solo sabe que tiene seis aminoácidos de longitud y que contiene una serina S, una treonina T, una cisteína C, una arginina R y dos glutamatos E. ¿Cuál es la probabilidad de que la secuencia "SECRET" ocurra por casualidad?
4. *i)*¿Cuántos números de 3 cifras puede formar con los dígitos 1,2,3,4 y 5? ¿Cuántos puede formar que no tengan ningún dígito repetido? *ii)*Suponga que tiene N objetos distintos. ¿Cuántos conjuntos distintos de m objetos puede formar? Si le resulta más fácil, considere un caso particular, por ejemplo $N = 5$, $m = 3$ y después generalice. *iii)*¿Qué significa la probabilidad condicional? *iv)*¿Qué significa que dos eventos son independientes entre sí? *v)*¿Qué es una densidad de probabilidad?
5. (*tomado del libro "Molecular Driving Forces"*) Considere un dado (de 6 caras) no cargado.
   - (a) ¿Cuántas veces tiene que arrojarlo para que la probabilidad de que salga el número 2 al menos una vez sea mayor o igual a 2/3?
   - (b)Suponga ahora que lo arroja 3 veces, ¿Cuál es la probabilidad de que el número 5 salga exactamente 2 de esas 3 veces?
   - (c) ¿Cuál es la probabilidad de que el 5 salga al menos dos de esas 3 veces?

6. (*tomado del libro "Molecular Driving Forces"*) Suponga que está comparando secuencias de aminoácidos de proteínas. Cada secuencia es una cadena de n letras de longitud. Suponga que tiene una secuencia de prueba a la que quiere comparar con s distintas secuencias tomadas de una bases de datos. Suponga que cualquiera de las letras del alfabeto (en este caso, los 20 aminoácidos) puede ser encontrada en cualquier posición de la secuencia, independientemente de qué letras están ubicadas en otras posiciones. Sea p la probabilidad de que la secuencia de prueba y una de las de la base de datos tengan el mismo aminoácido en una dada posición (cualquiera sea dicha posición).

   - (a) Calcule en términos de n, p y s el valor medio del número de coincidencias perfectas (el valor medio del número de veces que, en s comparaciones, la secuencia de prueba y la de la base de datos están caracterizadas por la misma secuencia ordenada de letras).
   - (b) ¿Cuál es el valor medio del número de veces en que, en s comparaciones, la secuencia de prueba y la de la base de datos difieren en un único aminoácido en una posición cualquiera?

7. Suponga que tira una moneda N veces seguidas y que, en cada intento, la probabilidad de que salga cara es p. *i)*¿Cuál es la probabilidad de que salga una secuencia determinada que tiene m caras y $N-m$ cecas? *ii)*¿Cuál es la probabilidad de que salgan exactamente m caras (es decir, que salga alguna secuencia con m caras y $N-m$ cecas)? Esta es la distribución binomial. *iii)*¿Cuál es la probabilidad de que salgan al menos m caras? *iv)*Si repite el experimento varias veces, ¿cuál es el valor medio del número de caras que espera obtener? *v)*Conteste todas las preguntas anteriores para el caso particular $p = 1/2$ y diga qué significan los resultados obtenidos. *vi)* Suponga que $p \ll 1$ y $N \gg 1$ de modo que $pN \approx \alpha$, ¿cómo puede aproximar la probabilidad de que salgan m caras? ¿Qué significado tiene α en términos del valor medio obtenido en el punto *iv)*? Esta aproximación corresponde a la distribución de Poisson.

8. (*tomado del libro "Molecular Driving Forces"*) De acuerdo a la teoría cinética de los gases, la distribución de velocidades, v_x, a lo largo de la dirección espacial, x, de un gas de moléculas de masa, m, está dada por la ley de Maxwell-Boltzmann, $f(v_x) = \alpha \exp(-mv_x^2/2k_BT)$ donde k_B es la constante de Boltzmann, T es la temperatura y α es un factor de normalización. $f(v_x)$ es una densidad de probabilidad en el espacio de velocidades, v_x. Las velocidades en esta descripción pueden tomar cualquier valor entre $-\infty$ e ∞.

   - (a) A partir de qué condición sobre la densidad de probabilidad se puede determinar α (el valor que se obtiene es α).
   - (b) Calcule la energía promedio $\langle mv_x^2/2\rangle$
   - (c) Calcule la velocidad promedio, $\langle v_x\rangle$.

9. (*tomado del libro "Molecular Driving Forces"*) Considere la distribución de probabilidad $p(x) = ax^n$, $0 \leq x \leq 1$, con n un número entero positivo.

   - (a) Obtenga la expresión de la constante a que normaliza a $p(x)$.
   - (b) Calcule el valor medio (o valor esperado) $\langle x\rangle$ en función de n.
   - (c) Calcule la varianza, $\sigma^2 = \langle x^2\rangle$ - $\langle x\rangle^2$ en función de n.

10. (*tomado del libro "Physical Biology of the Cell"*) Considere la estimación de variabilidad inter-celular hecha en las pp. 44-46 del libro que se incluyen al final de la guía. De acuerdo a ese cálculo el desvío estándar en el número de moléculas que va a una de dos células hijas durante la división celular viene es:

$$\langle n_1^2 \rangle - \langle n_1 \rangle^2 = Npq. \tag{1}$$

- (a) Obtenga este resultado.
- (b) Derive el resultado de que la diferencia promedio en la intensidad entre las dos células hijas viene dada por

$$\langle (I_1 - I_2)^2 \rangle = \alpha I_{tot}, \tag{2}$$

  donde I_1 e I_2 son las intensidades de las hijas 1 y 2, respectivamente, e I_{tot} es la intensidad de fluorescencia total de la célula madre y donde se supuso que existe una relación lineal entre la intensidad y el número de fluoróforos de la forma $I = \alpha N$.

11. *Simulación numérica del curso de Rob Phillips.* Hagan las simulaciones sobre tirar monedas de: `http://www.rpgroup.caltech.edu/aph161/assets/tut/t2/binomial_partitioning.html`. Modifiquen el código usando otros valores de la probabilidad, p. Reescriban el código para el caso de un dado de 6 caras que no está cargado.

12. *Simulación numérica (tomada del libro "Physical Biology of the Cell")* El simple resultado analítico de la Ecuación (2) se puede apreciar mejor con una simulación estocástica. La idea es "simular" la dilución de proteínas fluorescentes durante divisiones celulares repetidas donde se supone que todas las células madre comienzan con N copias de la proteína fluorescente y la distribución de esas proteínas a las dos hijas se determina estrictamente mediante un simple proceso aleatorio binario (es decir, cara o ceca de una moneda lanzada al aire). Conceptualmente, el algoritmo para realizar esta simulación se puede resumir así:

- 1. Elija el número de moléculas fluorescentes N en la célula madre y calcule la intensidad $I_{tot} = \alpha N$.
- 2. Genere N números aleatorios entre 0 y 1. Si un número aleatorio dado está entre 0 y 0,5, asigne la proteína a la hija 1 y, de lo contrario, asigne la proteína a la hija 2.
- 3. Calcule la "intensidad" de las hijas multiplicando N_1 y N_2 por α.
- 4. Realice la misma operación para 100 células (por ejemplo).
- 5. Elija una nueva N y repita el proceso.

Al considerar el mismo proceso de división una y otra vez, para diferentes elecciones de N, podemos generar una curva que capture la estadística establecida anteriormente. En particular, para cada N, se simularon 100 "divisiones" separadas y los resultados de cada proceso de división se muestran como puntos de datos. Para un valor dado de N, podemos calcular la raíz del error cuadrático medio en la partición y su intensidad correspondiente, dada por $\sqrt{\langle (I_1 - I_2)^2 \rangle}$. Para ver un ejemplo del resultado de tal simulación, examine la Figura 2.11 de las páginas del libro incluidas más abajo.

13. *Simulación numérica del curso de Rob Phillips.* Hagan las simulaciones sobre paseo al azar del "Ejemplo I" de: `http://www.rpgroup.caltech.edu/aph161/assets/tut/t3/t03_stochastic_simulations.html`. Agreguen el gráfico de un histograma adicional correspondiente a un tiempo intermedio entre

los dos graficados. Modifiquen el código para que haya una probabilidad no nula de que el paseador no se desplace en un paso temporal. Modifiquen el código de modo que sea más probable que el paseador se desplace en un sentido que en el otro. Comparen todos los histogramas.