Modelos para variables aleatorias discretas en R

1 Introducción: variables aleatorias discretas

Ejercicio 1.1

La variable aleatoria X representa los posibles resultados que se pueden obtener cuando se lanza al aire un dado de seis caras y la variable aleatoria Y representa la suma de los resultados obtenidos al lanzar el mismo dado dos veces. Asumiendo que el dado no está trucado, y que al lanzarlo, todos los resultados son equiprobables, obtenga: la función de masa de probabilidad de la variable aleatoria X y de la variable aleatoria Y, los valores esperados (o la esperanza matemática) en ambos casos E[X] y E[Y], los valores esperados de cada una de las dos variables aleatorias consideradas al cuadrado E[X^2] y E[Y^2] y, finalmente, las varianzas V[X] y V[Y].

Instrucciones:

1. Obtenga la función masa de probabilidad de la variable aleatoria X utilizando la función *rep (),* que tiene la siguiente sintaxis: *rep* (*p*, *times* = *n*).
2. Obtenga la función masa de probabilidad de la variable aleatoria Y. Para ello, cree en primer lugar un vector (llámelo X) que contenga todos los posibles resultados asociados a la variable aleatoria X y posteriormente, a partir de dicho vector, cree una tabla cuyas entradas sean las sumas de todos los posibles resultados que se pueden obtener al lanzar el dado dos veces, es decir, una tabla que represente todos los posibles resultados de la variable aleatoria Y. Utilice para ello la función *outer* () que tiene la siguiente sintaxis: *outer* (*X, X, FUN = '+'*). A partir de la tabla anterior cree una tabla de frecuencias y finalmente, a partir de ésta, obtenga la función de masa de probabilidad de Y pedida.
3. Obtenga las esperanzas matemáticas pedidas ayudándose para ello de la función *sum* ().
4. Calcule las varianzas pedidas a partir de los resultados de los apartados anteriores.

Ejercicio 1.2

En cierta tribu del sur del Amazonas, las familias suelen tener de 3 a 6 hijos. La función masa de probabilidad estimada de la variable aleatoria asociada X, es la siguiente: P(X = 3) = 0.35, P(X = 4) = 0.45, P(X = 5) = 0.16, P(X = 6) = 0.04. A partir de esta información, calcule E [X], V [X], y obtenga además la función de distribución de probabilidad acumulada correspondiente, F(x). Posteriormente, determine cuál es, para estas familias, la probabilidad de tener no más de 4 hijos y cuál es la probabilidad de tener más de 3 pero menos de 6 hijos.

Instrucciones:

1. Cree el vector que representa el espacio muestral de *X*.
2. Calcule E [*X*] teniendo en cuenta la función de masa de probabilidad del enunciado.
3. Obtenga la función de distribución de probabilidad acumulada utilizando para ello la función *cumsum* ().
4. Calcule las probabilidades pedidas a partir de los resultados anteriores.

2 Distribución binomial

Ejercicio 2.1

Asumiendo que la determinación del género un bebé durante su gestación sigue una distribución binomial, obtenga la función de masa de probabilidad y la función de densidad de probabilidad cuando la variable de interés es el número de hijas en una familia que tenga en total 5 hij@s, y represéntelas gráficamente.

Instrucciones:

1. Calcule en primer lugar la función de masa de probabilidad utilizando la función *dbinom* () cuya sintaxis es la siguiente: *dbinom* (*x*, *n*, *p* = *probability*) donde *x* = número de éxitos, *n* = número de experimentos y *p* = probabilidad de éxito.
2. Calcule posteriormente la función de distribución de probabilidad acumulada utilizando la función *pbinom* ().
3. Represente ambas funciones gráficamente utilizando la función *plot* ().

Ejercicio 2.2

Una determinada especie de ratón desarrolla una forma particular de distrofia muscular que tiene una clara base genética. Se sabe que, en este grupo, la probabilidad de aparición de la citada distrofia muscular es de ¼. Si se analizase una muestra de un total de 20 ratones de esta especie, calcule las probabilidades de que: (a) menos de cinco tuviesen distrofia muscular, (b) justo cinco tuviesen distrofia muscular, (c) más de dos pero menos de 8 tuvieran distrofia muscular.

Instrucciones:

1. Determine los parámetros del modelo de variable aleatoria a utilizar.
2. Calcule las probabilidades pedidas (a), (b) y (c) utilizando las funciones *pbinom* () y *dbinom* ().

Ejercicio 2.3

Suponga que la probabilidad de que un paciente que ha contraído ébola se recupere es de 0.1. Si 16 personas (sin relación de parentesco entre ellas) han contraído el virus, calcule: (a) el número esperado de pacientes que se recuperarán, (b) la probabilidad de que como mucho 5 se recuperen, (c) la probabilidad de que al menos 5 se recuperen, (d) la probabilidad de que exactamente 5 se recuperen.

Instrucciones:

1. Determine los parámetros del modelo de variable aleatoria a utilizar.
2. Determine el número esperado de pacientes que se recuperarán (a) utilizando la fórmula adecuado según el modelo utilizado.
3. Calcule las probabilidades pedidas (b), (c) y (d) utilizando las funciones *pbinom* () y *dbinom* ().

   NOTA: Investigue cómo resolver en R el apartado (c) de dos formas distintas.

Ejercicio 2.4

Imagine que una de las máquinas expendedoras de tickets del aparcamiento de cierto hospital está averiada y no siempre genera el ticket correspondiente. Asumiendo n = 181 y p = 0.15 (siendo p la probabilidad de éxito, es decir, la probabilidad de que la máquina genere el ticket), obtenga la gráfica de la función masa de probabilidad de la variable binomial $X \sim B(n,p)$ asociada a este problema.

Instrucciones:

1. Defina n, p, y un vector que contenga todos los posibles valores de X.
2. Utilice *dbinom* () para generar una variable que represente el valor de la función de masa de probabilidad en función de X.
3. Represente en R dicha función de masa de probabilidad utilizando *plot* ().

3 Distribución geométrica

Ejercicio 3.1

Una máquina que fabrica piezas para prótesis de rodilla tiene una probabilidad de fallo del 3% (p = 0.03) Asumiendo que esta probabilidad de fallo se puede considerar constante, se pide calcular una serie de probabilidades según las instrucciones que se adjuntan a continuación.

Instrucciones:

1. Calcule la probabilidad de que la máquina falle en la quinta pieza fabricada utilizando *dgeom* ().
2. Calcule la probabilidad de que el primer fallo ocurra antes de fabricar la sexta pieza.
3. Represente en R gráficamente la función de masa de probabilidad de la variable aleatoria geométrica asociada a este problema.
4. ¿Qué ocurre en la gráfica de la función de masa de probabilidad si se modifica el valor de *p*? ¿Y con la función de distribución de probabilidad? Represente ambas gráficas para tres valores de *p* distintos y comente las diferencias. Utilice *par* (*mfrow* = *c* (*number_of_rows*, *number_of_columns*)) si quiere representar varias gráficas a la vez.

NOTA: Tenga en cuenta que en R la distribución geométrica cuenta el número de experimentos sin incluir el último, que sería el primer éxito.

Ejercicio 3.2

Se está probando un nuevo medicamento contra la alergia primaveral que, de momento, tiene una probabilidad de no funcionar del 20%. ¿Cuál sería el número máximo de pacientes con alergia que deberían tomar el medicamento para que la probabilidad de encontrar a uno en el que el medicamento no haya funcionado sea del 75%?

Instrucciones:

1. Plantee a mano el ejercicio.
2. Busque qué función relacionada con la distribución geométrica puede utilizar en R para resolver el ejercicio y obtenga con ella el resultado pedido.

Ejercicio 3.3

En el servicio de cirugía de un hospital, están probando nuevos procedimientos para realizar implantes cocleares en recién nacidos. Utilizando una distribución geométrica, resuelva los apartados siguientes.

Instrucciones:

1. Calcule la probabilidad de obtener dos fallos antes de obtener un éxito, si la probabilidad de éxito de uno de los nuevos procedimientos es del 70%.
2. Calcule la probabilidad de obtener como mucho 7 fallos antes de obtener un éxito, si en este caso, el procedimiento utilizado tiene una probabilidad de éxito del 30%.
3. Obtenga el número máximo de fallos que habría antes de un éxito en el 95% de los casos, si la probabilidad de éxito en esta ocasión es del 40%.

4 Distribución binomial negativa

Ejercicio 4.1

Para optar a una beca universitaria, los alumnos de ingeniería biomédica de cierta universidad deben haber aprobado al menos 10 asignaturas con al menos una calificación de 8.0. Si la probabilidad p de sacar al menos un 8.0 en cada una de estas asignaturas (bajo condición de independencia) es de un 40%, calcule la probabilidad que tiene un estudiante de poder optar a esta beca cuando lleve cursadas 15 asignaturas.

Instrucciones:

1. Plantee a mano el ejercicio.
2. Resuelva el ejercicio utilizando *dnbinom* () y teniendo en cuenta que en R la distribución binomial negativa cuenta el número de fallos hasta conseguir r éxitos.
3. Represente en R gráficamente la evolución de la función de masa de probabilidad de este modelo binomial negativo en función de los valores de X. ¿Cómo cambiaría esta gráfica para distintos valores de p?

5 Distribución hipergeométrica

Ejercicio 5.1

Imagine que en un colegio de educación especial estudian un total de 46 niñas y 49 niños. Una vez al año, se escoge al azar a 20 de est@s niñ@s para que participen en una obra de teatro que luego es representada para sus padres. ¿Cuál es la probabilidad de que entre l@s 20 niñ@s escogidos haya exactamente 5 niñas y 15 niños? ¿Y la probabilidad de que el número de niñas elegidas sea de cómo mucho 5?

Instrucciones:

1. Plantee a mano el ejercicio y obtenga la solución a la primera pregunta planteada pedida en R utilizando *choose* ().
2. Conteste a las dos preguntas planteadas resolviendo en este caso el ejercicio en R utilizando *dhyper* () o *phyper* () según corresponda.
3. Represente en R gráficamente tanto la función de masa de probabilidad como la función de distribución del modelo hipergeométrico utilizado en ese ejercicio cuando éste cuenta el número de niñas seleccionadas.

Ejercicio 5.2

En cierto hospital madrileño, hay 5 vacantes en el servicio de cirugía. De entre l@s candidat@s, 500 mujeres y 450 hombres cumplen los requisitos necesarios para el puesto, por lo que la elección final se va a realizar al azar. Bajo estas premisas, calcule las probabilidades pedidas según las instrucciones que se adjuntan a continuación.

Instrucciones:

1. Calcule la probabilidad de que de las 5 vacantes exactamente 4 sean ocupadas por mujeres utilizando un modelo hipergeométrico.
2. Calcule, con el mismo modelo, la probabilidad de que al menos 3 vacantes sean ocupadas por hombres.
3. Según las condiciones del problema, ¿es posible realizar una aproximación utilizando un modelo binomial? Si es así, utilícelo para calcular de nuevo las probabilidades pedidas en los dos apartados anteriores y compare los resultados obtenidos en ambos casos
4. Represente gráficamente las funciones de masa de probabilidad tanto del modelo binomial como del modelo geométrico utilizados en este problema.

6 Distribución de Poisson

Ejercicio 6.1

Se sabe que, en promedio, la llegada de pacientes al servicio de urgencias de cierto centro de salud, en un día, es de 2. Bajo estas premisas, siga las instrucciones para responder a las preguntas que se plantean a continuación.

Instrucciones:

1. Calcule la probabilidad de que en un día concreto no acuda ningún paciente al servicio de urgencias de dicho centro de salud. Utilice para ello la función *dpois* ().
2. Calcule la probabilidad de que en un día concreto lleguen menos de 6 pacientes al servicio de urgencias de este centro de salud. Ayúdese en R de la función *ppois* ().
3. Finalmente, obtenga la probabilidad de que en un día concreto la llegada de pacientes sea mayor que 3 pero menor que 6.

Ejercicio 6.2

Un grupo de estudiantes de ingeniería forestal, están realizando un estudio en áreas de 10 m x 10 m de bosque subtropical sobre la probabilidad de encontrar un número determinado de árboles. Bajo la premisa de que los árboles se encuentran distribuidos de manera aleatoria y teniendo en cuenta que la media de árboles en áreas de este tamaño es de 30, se pide calcular una serie de probabilidades que se describen a continuación.

Instrucciones:

1. Calcule la probabilidad de encontrar no más de 3 árboles en un área de 1 m^2 ajustando en primer lugar el valor de la media a la nueva área bajo estudio y utilizando posteriormente la función *ppois* ().
2. Calcule la probabilidad de encontrar exactamente 1 árbol en un área de 5 m^2 utilizando la función *dpois* ().
3. Finalmente, calcule la probabilidad de encontrar al menos 2 árboles en un área de 2 m^2. Intente realizar este último cálculo en R de dos maneras distintas. Explore para ello las posibilidades de la función *ppois* ().

Ejercicio 6.3

Cierta enfermedad, que suele ser detectada justo al nacer, tiene una prevalencia (p) del 0.01%. Si se realiza un estudio acerca de esta enfermedad en un hospital en el que se sabe que han nacido n = 5000 niños en el último año, calcule las probabilidades pedidas según las instrucciones que se adjuntan a continuación.

Instrucciones:

1. Calcule cuál es la probabilidad de que, de entre esos 5000 niños, haya al menos uno con la citada enfermedad utilizando en primer lugar un modelo binomial.
2. Justifique si es posible o no utilizar como aproximación un modelo de Poisson, y en caso afirmativo, resuelva de nuevo el apartado anterior y compare los resultados obtenidos.
3. Calcule la probabilidad de que no haya más de dos niños con la enfermedad de entre los 5000 considerados utilizando un modelo binomial.
4. De nuevo, si es posible utilizar un modelo de Poisson, resuelva de nuevo el apartado anterior y compare los resultados obtenidos.
5. Represente sobre la misma gráfica, la función de masa de probabilidad de los dos modelos utilizados en el ejercicio y comente las similitudes y/o diferencias entre ambas.
6. ¿Qué ocurre en la comparativa del modelo binomial con el modelo de Poisson si, dejando constante el producto $n \cdot p$, se disminuye n y se aumenta p? Represente gráficamente las funciones de masa de probabilidad para distintas parejas (n,p) y comente el resultado.