OCW-UPM Estadística para Ingeniería Civil y Medioambiental

Autores: E. M. García del Toro, C. Hermoso, E. J. Huertas

PROBLEMAS Y EJERCICIOS RESUELTOS

TEMA 2 - REGRESIÓN

Ejercicios resueltos de álgebra y probabilidad de sucesos

Ejercicios de repaso y distribuciones discretas en general

Ejercicios resueltos de distribución binomial

Ejercicio 1.- Una empresa química produce barras de grafito de elevadísima pureza para su uso en centrales nucleares. Su fabricación es tan delicada que, como promedio general, el 30 % de ellas salen defectuosas y hay que desecharlas. Las barras se fabrican una a una, y de forma completamente aislada unas de otras. Calcular las siguientes probabilidades para el próximo lote de fabricación de 50 barras.

a) ¿Qué distribución de probabilidad tiene la variable aleatoria $X \equiv$ "*número de barras defectuosas de entre las próximas 50 fabricadas*"?

b) ¿Cuál es la probabilidad de encontrar exactamente 3 barras defectuosas?

c) ¿Y la de encontrar más de 2 barras defectuosas?

d) ¿Cuántas barras defectuosas se espera encontrar en las 50 del próximo lote? Justifique la respuesta.

Para resolver problemas de este tipo lo primero es darse cuenta que la variable aleatoria necesaria para resolverlos se distribuye como una Binomial $B(n,p)$ de parámetros n y p. En el enunciado siempre deben proporcionar la información para obtener ambos.

Solución ejercicio 1:

Apartado a) En el enunciado ya nos están diciendo quien debe ser la variable aleatoria discreta necesaria para responder a las preguntas (esto no ocurre en todos los problemas de este tipo). En este caso es $X \equiv$ "*número de barras defectuosas de entre las próximas 50 fabricadas*". Los únicos valores posibles que puede tomar la variable X son $0, 1, 2, \ldots 50$. Es decir, el **soporte** de X son los siguientes 51 valores (démonos cuenta que aunque $n = 50$, el número de valores del soporte son en este caso 51)

$$S_X = \{0, 1, 2, \ldots 50\}$$

Ahora analicemos la forma en la que se lleva a cabo el estudio del que nos habla el problema. Una vez fabricadas las 50 barras, se mira cada una de ellas, y si está defectuosa se considera un éxito, ya que estamos contando barras defectuosas. Si por contra la barra es correcta, se considera un fracaso. La probabilidad de éxito p viene dada en el enunciado, donde dicen que en general salen defectuosas el 30 % de las barras fabricadas, lo que implica una probabilidad de éxito $p = 0.3$. El examen **de cada una de las barras** por separado, es en realidad un proceso de Bernoulli independiente, con probabilidad de éxito $p = 0.3$ y probabilidad de fracaso $q = 1 - p = 0.7$. Como se van a analizar 50 barras de grafito, la variable X se distribuye como una **binomial** de parámetros $n = 50$ y $p = 0.3$, ya que lo que vamos a hacer con X es **contar** el número de éxitos en una secuencia de $n = 50$ ensayos de Bernoulli independientes entre sí, con una probabilidad fija $p = 0.3$ de ocurrencia del éxito entre los diferentes ensayos. Este hecho se escribe de la siguiente forma

$$X \sim B(n = 50; p = 0.3)$$

Apartado b) Ahora utilizamos la fórmula que otorga en este caso las correspondientes probabilidades a cada uno de los puntos del soporte de X. Esta es la **función de probabilidad** para una distribución binomial

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \text{ con } k = 0, 1, 2, \ldots 50. \tag{1}$$

Así, tenemos que la probabilidad de encontrar exactamente 3 barras defectuosas es

$$P(X = 3) = \binom{50}{3} \cdot 0.3^3 \cdot 0.7^{50-3} = 0.00002775,$$

ya que

$$\binom{50}{3} = \frac{50!}{3!(50-3)!} = \frac{50 \cdot 49 \cdot 48 \cdot 47!}{3! \cdot 47!} = \frac{50 \cdot 49 \cdot 48}{3 \cdot 2} = 19600,$$

$0.3^3 = 0.027$, y $0.7^{50-3} = 0.7^{47} = 5.2433 \times 10^{-8}$. El producto de los valores anteriores da el resultado pedido. Observamos que es una probabilidad muy baja, ya que estoy pidiendo que sean **exactamente** tres las barras defectuosas que me encuentro entre las 50 del próximo lote.

Apartado c) La probabilidad que nos piden en este apartado va a ser mucho mayor, ya que ahora me piden la probabilidad de encontrar **más de 2 barras defectuosas**, por lo que el numero de sucesos que satisfacen esto es mucho mayor. Puedo encontrar 3, 4, 5, 6, etc, etc, hasta 50. Así pues

$$P(X > 2) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + \cdots + P(X = 50).$$

Si realizamos la suma anterior, tendremos que calcular gran cantidad de probabilidades individuales, por lo que si calculamos una a una tardaremos muchísimo tiempo. La clave aquí está en darse cuenta de que la **probabilidad total** repartida por (1) entre los 51 valores posibles que puede tomar X **debe ser exactamente igual a uno**.

$$\sum_{k=0}^{50} P(X = k) = \sum_{k=0}^{50} \binom{50}{k} 0.3^k 0.7^{50-k} = 1$$

En la Figura 1 podemos ver la gráfica de la **función de probabilidad** $P(X = k)$ del problema, frente a los 51 posibles valores de $k = 0, 1, 2 \ldots 50$.

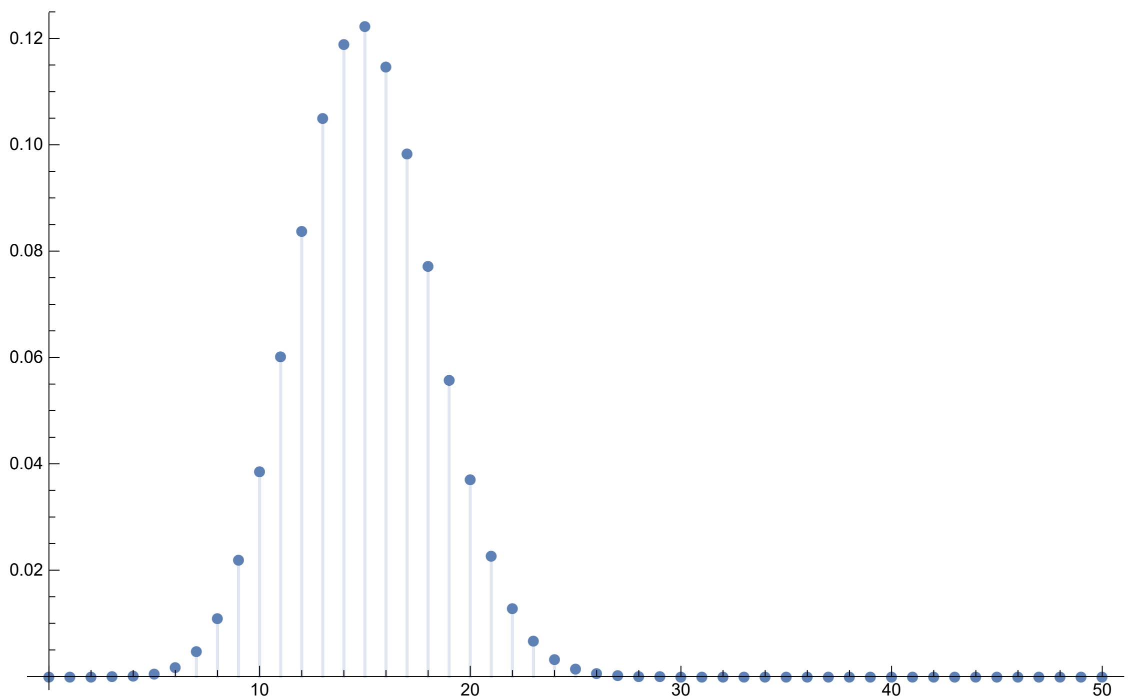

Figure 1: Función de probabilidad del Ejemplo 1

Sabiendo esto, podemos calcular la probabilidad $P(X > 2)$ de la siguiente forma

$$P(X > 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

y así solo tenemos que calcular tres probabilidades individuales y restárselas a la probabilidad total 1. Utilizando (1) de nuevo tenemos

$$P(X = 0) = \binom{50}{0} \cdot 0.3^0 \cdot 0.7^{50-0} = 1.7985 \times 10^{-8},$$

$$P(X = 1) = \binom{50}{1} \cdot 0.3^1 \cdot 0.7^{50-1} = 3.854 \times 10^{-7},$$

$$P(X=2)=\binom{50}{2}\cdot 0.3^2\cdot 0.7^{50-2}=4.047\times 10^{-6}.$$

Luego
$$P(X>2)=1-\left[1.7985\times 10^{-8}+3.8539\times 10^{-7}+4.047\times 10^{-6}\right]=0,999996$$
Esta probabilidad es prácticamente uno. Esto indica que el hecho de encontrar más de dos barras defectuosas entre las próximas 50 fabricadas es prácticamente seguro.

Apartado d) El número esperado de barras defectuosas será simplemente la esperanza, o valor esperado para la variable aleatoria X, que es la variable que las está "contando". Para la distribución binomial el valor esperado es siempre $\mu=E[X]=n\cdot p$, que en nuestro caso toma el valor

$$\mu=50\cdot 0.3=15 \text{ barras.}$$

Fijémonos que el máximo de probabilidad en la gráfica de la Figura 1 de la función de probabilidad $P(X=k)$ de este problema está precisamente en el valor $k=15$, luego este es el **valor más probable** que puede tomar X de entre todos los de su soporte.

Ejercicio 2 - Una asociación de cardiología afirma que solo el 10 % de los adultos mayores de 30 años logran completar una prueba de esfuerzo físico especialmente diseñada para ellos. Se toman al azar cuatro personas mayores de 30 años y se someten a una prueba de esfuerzo. Calcule la probabilidad de que:
a) Dos personas pasen la prueba.
b) Ninguna persona pase la prueba.
c) Una sola persona no pase la prueba.
d) Más de una, pero menos de cuatro personas pasen la prueba.
e) Una o más, pero tres o menos, no pasen la prueba.
f) Si se toma una muestra de 50 personas, ¿cuál es la probabilidad de que al menos dos personas pasen la prueba.

El estudio cardiológico se lleva a cabo para solamente 4 personas mayores de 30 años. Una persona se somete a la prueba y la pasa (éxito) o no la pasa (fracaso). La probabilidad de éxito p viene dada en el enunciado, ya que nos dicen que solo el 10 % de los adultos mayores de 30 años pasan la prueba, luego esto implica una $p=0.1$. Como hay 4 adultos mayores de 30 años que se someten a la prueba de esfuerzo, necesitamos una variable aleatoria X que cuente el número de personas mayores de 30 años que pasan la prueba de esfuerzo. La variable X se distribuye como una **binomial** de parámetros $n=4$ y $p=0.1$. En la Figura 2 podemos ver la gráfica de la **función de probabilidad** $P(X=k)$ de este segundo ejemplo, frente a los 5 posibles valores del soporte $k=0,1,2,3,4$.

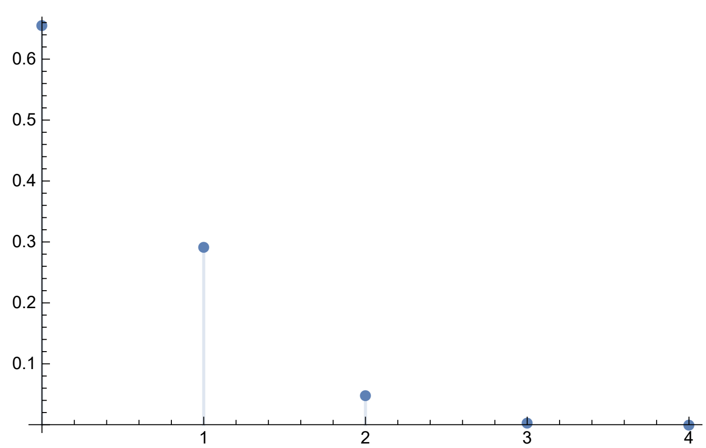

Figure 2: Función de probabilidad del Ejemplo 2

Solución ejercicio 2:
Como no dicen en el enunciado la variable aleatoria necesaria para resolver el problema, debemos proponerla nosotros. Queremos contar el número de personas mayores de 30 años, de entre las 4 que se realiza el estudio, que

pasan la prueba de esfuerzo. Así pues, tomamos $X \equiv$ "*número de personas mayores de 30 años, de entre las 4 que se realiza el estudio, que pasan la prueba de esfuerzo*". Los únicos valores posibles que puede tomar ahora la variable X son solamente los cinco valores $0, 1, 2, 3,$ y 4. El **soporte** de X es entonces $S_X = \{0, 1, 2, 3, 4\}$. Tenemos ahora

$$X \sim B(4, 0.1)$$

Apartado a) Nos preguntan $P(X = 2)$, luego utilizando nuevamente (1)

$$P(X = 2) = \binom{4}{2} \cdot 0.1^2 \cdot 0.9^{4-2} = 0.0486.$$

Apartado b) Nos preguntan $P(X = 0)$, luego

$$P(X = 0) = \binom{4}{0} \cdot 0.1^0 \cdot 0.9^{4-0} = 0.6561.$$

Apartado c) Nos preguntan $P(X = 3)$, ya que decir que una sola persona de entre las cuatro bajo estudio no pase la prueba es lo mismo que decir que exactamente tres de las cuatro personas pasan la prueba

$$P(X = 3) = \binom{4}{3} \cdot 0.1^3 \cdot 0.9^{4-3} = 0.0036.$$

Apartado d) Nos preguntan $P(1 < X < 3)$. Esto es simplemente sumas las probabilidades de los casos del soporte de X que se encuentran en este intervalo, es decir (valores que ya hemos calculado)

$$P(1 < X < 3) = P(X = 2) + P(X = 3) = 0.0486 + 0.0036 = 0.0522$$

Apartado e) El que una o más, pero tres o menos no pasen la prueba equivale a que la pasen cuatro o ninguno, luego nos piden la suma de las dos probabilidades siguientes

$$P(X = 4) + P(X = 0) = 0.0001 + 0.6561 = 0.6562.$$

Apartado f) Si ahora el estudio se amplia a 50 personas **debemos de cambiar de variable aleatoria**, ya que X solo contaba de 0 a 4 personas. Definimos la nueva variable aleatoria $Y \equiv$ "*número de personas mayores de 30 años, de entre las 50 que se realiza el estudio, que pasan la prueba de esfuerzo*". Esta nueva variable aleatoria se distribuye como

$$Y \sim B(50, 0.1)$$

y su funcion de probabilidad aparece en la Figura 3. Obviamente el soporte de Y es ahora

$$S_Y = \{0, 1, 2, \ldots 50\}.$$

En estas condiciones, nos preguntan $P(Y \geq 2)$ que podemos calcularla de la siguiente manera (**¡es muy importante entender esto!**)

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - 0.0338 = 0.9662$$

Ejercicio 3 - Cierto parque eólico de montaña se compone de un total de 43 aerogeneradores, suficientemente separados entre sí, y que operan independientemente unos de otros. En las especificaciones de cada aerogenerador se indica que la probabilidad de romperse frente a vientos de tipo huracanado es de $p = 0.05$. El parque acaba de sufrir un temporal de vientos huracanados, e interesa predecir los daños ocurridos allí para enviar un equipo técnico más o menos numeroso a repararlos. Calcular:

a) ¿Qué distribución de probabilidad tiene la variable aleatoria $X \equiv$ "*número de aerogeneradores rotos durante el temporal, de los 43 operativos en el parque*"? Indicar el tipo de distribución y el valor de sus correspondientes parámetros.

b) ¿Cuál es la probabilidad de que se no se haya roto ningún aerogenerador?

c) ¿Cuál es la probabilidad de que se hayan roto al menos dos aerogeneradores?

d) ¿Cuál es la probabilidad de que se hayan roto exactamente entre 3 y 5 aerogeneradores?

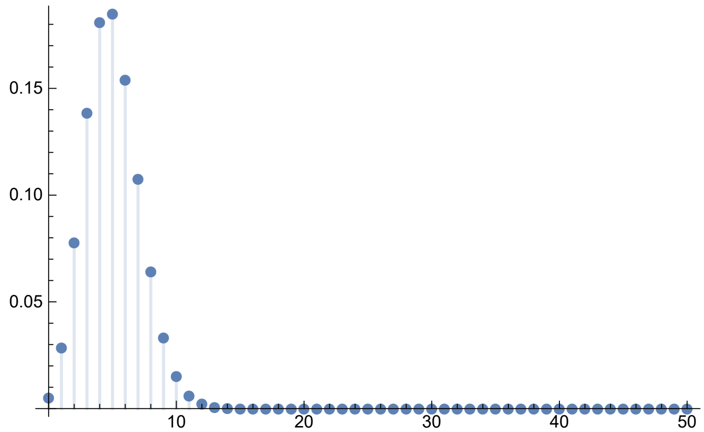

Figure 3: Función de probabilidad del apartado f) Ejemplo 2

e) ¿Cuál es el número esperado de aerogeneradores rotos que encontrará el equipo de reparación?

Solución ejercicio 3:

Apartado a) Tomamos como "éxito" el que se rompa un aerogenerador durante el temporal. La ruptura (o no) de un solo aerogenerador, se puede modelar mediante un proceso de Bernoulli independiente con probabilidad de "éxito" $p = 0.05$ y probabilidad de "fracaso"

$$q = 1 - 0.05 = 0.95.$$

La variable aleatoria discreta X se distribuye como una **binomial** de parámetros $n = 43$ y $p = 0.05$, ya que con ella **contamos** el número de "éxitos" en una secuencia de $n = 43$ ensayos de Bernoulli independientes entre sí, con una probabilidad fija $p = 0.05$ de ocurrencia de "éxito" entre los diferentes ensayos. Esto se escribe formalmente como

$$X \sim B(n = 43,\, p = 0.05).$$

Comentarios adicionales: El soporte de X (el conjunto completo de valores que puede tomar) tiene 44 **valores** (desde $X = 0$ hasta $X = 43$). La expresión que otorga las correspondientes probabilidades a cada uno de los puntos del soporte es la **función de probabilidad** para una distribución binomial

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \text{ con } k = 0, 1, 2, \ldots 43. \tag{2}$$

En la Figura 1 podemos ver la gráfica de la **función de probabilidad** $P\left(X = k\right)$ del problema

Apartado b) La probabilidad de que no se haya roto ningún generador durante el temporal es

$$P(X = 0) = \binom{43}{0} \cdot 0.05^0 \cdot 0.95^{43-0} = 0.11018.$$

Apartado c) La **probabilidad total** repartida por (2) entre los 44 valores posibles que puede tomar X **es exactamente igual a uno**.

$$\sum_{k=0}^{43} P(X = k) = \sum_{k=0}^{43} \binom{43}{k} \cdot 0.05^k \cdot 0.95^{43-k} = 1$$

El suceso $X \geq 2$ representa la ruptura de al menos dos aerogeneradores (es decir, 2 o más máquinas, hasta las 43 existentes) durante el temporal. Teniendo esto en cuenta,

$$\begin{aligned} P\left(X \geq 2\right) &= P\left(X = 2\right) + P\left(X = 3\right) + \cdots + P\left(X = 43\right) \\ &= 1 - \left[P\left(X = 0\right) + P\left(X = 1\right)\right]. \end{aligned}$$

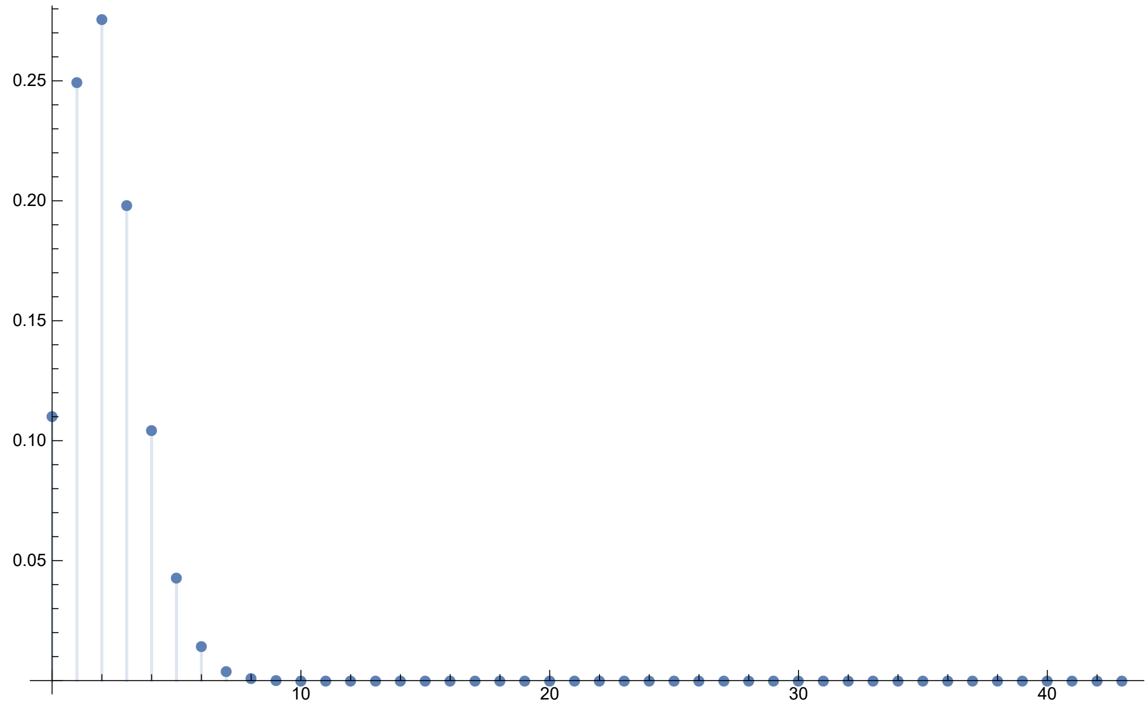

Figure 4: Función de probabilidad del problema

Conociendo de antes $P(X=0)=0.11018$ y calculando

$$P(X=1)=\binom{43}{1}\cdot 0.05^1\cdot 0.95^{43-1}=0.24936,$$

tenemos

$$P(X\geq 2)=1-[0.11018+0.24936]=0.64046.$$

Apartado d) La probabilidad pedida es la suma de tres probabilidades

$$P(X=3)=\binom{43}{3}\cdot 0.05^3\cdot 0.95^{43-3}=0.19825,$$

$$P(X=4)=\binom{43}{4}\cdot 0.05^4\cdot 0.95^{43-4}=0.10434,$$

$$P(X=5)=\binom{43}{5}\cdot 0.05^5\cdot 0.95^{43-5}=0.042834.$$

Luego

$$\begin{aligned} P(3\leq X\leq 5) &= P(X=3)+P(X=4)+P(X=5) \\ &= 0.19825+0.10434+0.042834 \\ &= 0.34542. \end{aligned}$$

Apartado d) El número pedido es el **valor esperado** de la variable aleatoria X. Para la distribución binomial el valor esperado es $E[X]=n\cdot p$, luego

$$E[X]=43\cdot 0.05=2.15 \text{ aerogeneradores.}$$

Ejercicio 4 - Se realiza un estudio sobre un nuevo tratamiento de la filoxera (*Dactylosphaera vitifoliae*) sobre 15 cepas de vid afectadas por el parásito. Se comprueba que el tratamiento **no** es efectivo en el 25 % de las cepas tratadas. Calcular:

a) Probabilidad de que se recupere al menos una de las cepas de vid tratadas.
b) Probabilidad de que se recupere a lo más una de las cepas de vid tratadas.

Solución ejercicio 4.

Para contar el número de cepas de vid que se recuperan después del tratamiento definimos la variable aleatoria discreta $X \equiv$ "*número de cepas de vid, de entre las 15 del estudio, que se recuperan después del tratamiento*". Esta variable se distribuye de la forma $X \sim B(n = 15, p = 0.75)$, ya que si la probabilidad de fracaso del tratamiento sobre una cepa afectada de filoxera es $q = 0.25$, entonces la probabilidad de éxito del tratamiento es $p = 0.75$.

Apartado a) Nos preguntan la probabilidad de que se recupere o bien una sola cepa o más (hasta 15, porque solo se estudian 15) luego nos piden

$$P(X \geq 1) = P(X = 1) + P(X = 2) + \cdots + P(X = 15) = 1 - P(X = 0).$$

Como

$$P(X = 0) = \binom{15}{0} 0.75^0 \cdot 0.25^{15-0} = 0.25^{15} = 9.3132 \times 10^{-10}$$

es una probabilidad bajísima, muy próxima a cero, encontramos que $P(X \geq 1) = 1 - 9.3132 \times 10^{-10} \simeq 1$ (prácticamente certeza absoluta)

Apartado b) Nos preguntan la probabilidad de que se recupere o bien ninguna cepa o bien solo una, es decir, $P(X \leq 1)$. Como $P(X \leq 1) = P(X = 0) + P(X = 1)$, si

$$P(X = 1) = \binom{15}{1} 0.75^1 \cdot 0.25^{15-1} = 4.191 \times 10^{-8}$$

entonces $P(X \leq 1) = P(X = 0) + P(X = 1) = 9.3132 \times 10^{-10} + 4.191 \times 10^{-8} = 4.2841 \times 10^{-8}$ que es prácticamente cero (es casi imposible que solo se recupere una o ninguna cepa).

Ejercicio 5 - Se realiza un ensayo sobre un nuevo insecticida y se comprueba que **no** funciona en el 15% de los casos. Si se aplica el insecticida a 20 campos de cultivo diferentes, calcular:
a) Probabilidad de que se eliminen los insectos en al menos uno de los campos de cultivo tratados.
b) Probabilidad de que se eliminen los insectos a lo más en uno de los campos de cultivo tratados.

Solución ejercicio 5.

$X \equiv$ "*número de campos de cultivo con insectos, de entre los 20 tratados con el insecticida, que se recuperan después del tratamiento*".

$$X \sim B(n = 20, p = 0.85)$$

Apartado a)

$$P(X \geq 1) = P(X = 1) + P(X = 2) + \cdots + P(X = 20) = 1 - P(X = 0).$$

$$P(X = 0) = \binom{20}{0} 0.85^0 \cdot 0.15^{20-0} = 0.25^{15} = 3.3253 \times 10^{-17}$$

$$P(X \geq 1) = 1 - 3.3253 \times 10^{-17} \simeq 1$$

Apartado b)

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$P(X = 1) = \binom{20}{1} 0.85^1 \cdot 0.15^{20-1} = 3.7686 \times 10^{-15}$$

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= 3.3253 \times 10^{-17} + 3.7686 \times 10^{-15} \\ &= 3.8019 \times 10^{-15} \simeq 0 \end{aligned}$$

Ejercicio 6 - Se realiza un estudio sobre un nuevo larvicida y se comprueba que **no** funciona en el 15% de los casos. Si se aplica el tratamiento a 20 plantas enfermas, calcular:

a) Probabilidad de que se recupere al menos una de las plantas tratadas.

b) Probabilidad de que se recupere a lo más una de las plantas tratadas.

Solución ejercicio 6

Para contar el número de plantas enfermas que se recuperan después del tratamiento con el larvicida, definimos la variable aleatoria discreta $X \equiv$ "*número de plantas enfermas, de entre las 20 tratadas con el larvicida, que se recuperan después del tratamiento*". Esta variable se distribuye de la forma $X \sim B(n = 20,\, p = 0.85)$, ya que si la probabilidad de fracaso del tratamiento sobre una planta enferma es $q = 0.15$, entonces la probabilidad de éxito del tratamiento es $p = 0.85$.

Apartado a) Nos preguntan la probabilidad de que se recupere o bien una sola planta o más (hasta 20, porque solo se tratan 20) luego nos piden

$$P(X \geq 1) = P(X = 1) + P(X = 2) + \cdots + P(X = 20) = 1 - P(X = 0).$$

Como

$$P(X = 0) = \binom{20}{0} 0.85^0 \cdot 0.15^{20-0} = 0.25^{15} = 3.3253 \times 10^{-17}$$

es una probabilidad extremadamente pequeña, muy próxima a cero, encontramos que $P(X \geq 1) = 1 - 3.3253 \times 10^{-17} \simeq 1$ (prácticamente certeza absoluta de que se recuperará al menos una planta)

Apartado b) Nos preguntan la probabilidad de que se recupere o bien ninguna planta o bien solo una, es decir, $P(X \leq 1)$. Como $P(X \leq 1) = P(X = 0) + P(X = 1)$, si

$$P(X = 1) = \binom{20}{1} 0.85^1 \cdot 0.15^{20-1} = 3.7686 \times 10^{-15}$$

entonces $P(X \leq 1) = P(X = 0) + P(X = 1) = 3.3253 \times 10^{-17} + 3.7686 \times 10^{-15} = 3.8019 \times 10^{-15}$ que es prácticamente cero (es casi imposible que solo se recupere una o ninguna planta enferma después del tratamiento).

Ejercicios resueltos de distribución de Poisson

Ejercicio 7 - Una central nuclear de construcción antigua sufre accidentes en forma de pequeños escapes radioactivos cada cierto tiempo. En promedio, se ha contabilizado 1 accidente cada 8 meses, y se supone que el número de escapes de la central sigue un proceso de Poisson. Determine las probabilidades siguientes:

a) Probabilidad de que no ocurra ningún escape durante el próximo mes.

b) Probabilidad de que no ocurra ningún escape durante el próximo año.

c) Probabilidad de que ocurra exactamente un escape en el próximo mes.

d) Probabilidad de que ocurran a lo más 3 escapes durante el próximo año.

e) Probabilidad de que ocurran 3 o más escapes durante el próximo año.

f) Probabilidad de que ocurran entre 2 y 5 escapes durante el próximo año.

g) Probabilidad de que ocurran entre 2 y 5 escapes durante los próximos tres años.

Los problemas de este tipo es conveniente resolverlos **apartado por apartado**, utilizando en cada apartado una variable aleatoria que cuente el **número de sucesos en el tiempo que nos indiquen**, teniendo siempre en cuenta el promedio que nos da el enunciado. En este caso, los sucesos son accidentes en una central nuclear.

Solución ejercicio 7:

Apartado a) Por ejemplo, para resolver este apartado consideramos la variable aleatoria $X_a \equiv$ "*número de escapes radioactivos en 1 mes*". A continuación debemos averiguar el λ_a correspondiente a este apartado, que usualmente se obtiene planteando una regla de tres simple con el promedio que proporciona el enunciado (1 escape cada 8 meses). Así,

$$\left\{ \begin{array}{lcl} 1 \text{ escape} & \rightarrow & 8 \text{ meses} \\ \lambda_a \text{ escapes} & \leftarrow & 1 \text{ mes} \end{array} \right. ,$$

luego

$$\lambda_a = \frac{1 \times 1}{8} \frac{\text{escapes}}{\text{mes}} = 0.125 \frac{\text{escapes}}{\text{mes}}.$$

Asumimos que X_a se distribuye como una Poisson de parámetro $\lambda_a = 0.125$, es decir $X_a \sim Poisson(\lambda_a = 0.125)$, y utilizando la fórmula para la función de probabilidad de la distribución de Poisson

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{3}$$

tenemos que la probabilidad de sufrir 0 accidentes en un mes es $P(X_a = 0 \text{ escapes})$, i.e.,

$$P(X_a = 0) = \frac{e^{-\lambda_a}\lambda_a^0}{0!} = \frac{e^{-0.125} \times 0.125^0}{0!} = e^{-0.125} = 0.8825.$$

Apartado b) Ahora debemos establecer una nueva variable aleatoria X_b, ya que nos estamos refiriendo a otra cantidad de tiempo. Así pues, consideramos $X_b \equiv$ *"número de escapes radioactivos en 1 año"*. Igual que antes, tenemos que averiguar el λ_b correspondiente al promedio de escapes por año, que también obtenemos planteando una regla de tres simple con el promedio del enunciado. Así,

$$\begin{cases} 1 \text{ escape} & \rightarrow & 8 \text{ meses} \\ \lambda_b \text{ escapes} & \leftarrow & 1 \text{ año (12 meses)} \end{cases},$$

luego

$$\lambda_b = \frac{1 \times 12}{8} \frac{\text{escapes}}{\text{mes}} = 1.5 \frac{\text{escapes}}{12 \text{ meses}} = 1.5 \frac{\text{escapes}}{\text{año}}.$$

Asumimos ahora que X_b se distribuye como una Poisson de parámetro $\lambda_b = 1.5$, es decir $X_b \sim Poisson(\lambda_b = 1.5)$, y utilizando la fórmula para la función de probabilidad de la distribución de Poisson (3) tenemos que la probabilidad de sufrir 0 accidentes en un año es $P(X_b = 0 \text{ escapes})$, i.e.,

$$P(X_b = 0) = \frac{e^{-\lambda_b}\lambda_b^0}{0!} = \frac{e^{-1.5} \times 1.5^0}{0!} = e^{-1.5} = 0.22313.$$

Obviamente la probabilidad es bastante menor, porque es menos probable que pase más tiempo sin que haya un accidente.

Apartado c) Como ahora vuelven a preguntar por el número de escapes **en 1 mes**, volvemos a retomar la variable aleatoria del apartado a) y la misma λ_a. Ahora simplemente preguntan por la probabilidad de que $X_a = 1$ escape al mes, luego

$$P(X_a = 1) = \frac{e^{-\lambda_a}\lambda_a^1}{1!} = \frac{e^{-0.125} \times 0.125^1}{1!} = e^{-0.125} \times 0.125 = 0.11031.$$

Apartado d) Este apartado se refiere ahora al periodo de tiempo de un año nuevamente, por lo que volvemos a utilizar la variable aleatoria del apartado b) y la correspondiente λ_b. Así, la probabilidad de que ocurran a lo más 3 escapes durante el próximo año será la suma de las probabilidades de que ocurran entre 0 y 3 escapes el próximo año, ambas incluídas. Así

$$P(0 \leq X_b \leq 3) = P(X_b = 0) + P(X_b = 1) + P(X_b = 2) + P(X_b = 3).$$

La $P(X_b = 0) = 0.22313$, como vimos en el apartado b). Y además

$$\begin{aligned} P(X_b &= 1) = \frac{e^{-\lambda_b}\lambda_b^1}{1!} = \frac{e^{-1.5} \times 1.5^1}{1!} = 0.33470, \\ P(X_b &= 2) = \frac{e^{-\lambda_b}\lambda_b^2}{2!} = \frac{e^{-1.5} \times 1.5^2}{2!} = 0.25102, \\ P(X_b &= 3) = \frac{e^{-\lambda_b}\lambda_b^3}{3!} = \frac{e^{-1.5} \times 1.5^3}{3!} = 0.12551. \end{aligned}$$

Luego la probabilidad de que ocurran a lo más 3 escapes durante el próximo año es

$$P(0 \leq X_b \leq 3) = 0.22313 + 0.33470 + 0.25102 + 0.12551 = 0.93436$$

Apartado e) Continuamos con la misma variable aleatoria X_b referida a escapes durante 1 año. La probabilidad de que ocurran 3 o más escapes durante el próximo año es igual a la suma de todas las probabilidades para tres o más casos

$$P(X_b \geq 3) = \sum_{k=3}^{\infty} P(X_b = k)$$

que tiene un número infinito de sumandos. Como

$$\sum_{k=0}^{\infty} P(X=k) = 1$$

para cualquier variable aleatoria X que se distribuya como una Poisson, podemos calcular $P(X_b \geq 3)$ simplemente de la forma

$$P(X_b \geq 3) = 1 - P(X_b < 3) = 1 - P(0 \leq X_b \leq 2).$$

Como

$$\begin{aligned} P(0 &\leq X_b \leq 2) = P(X_b = 0) + P(X_b = 1) + P(X_b = 2) \\ &= 0.22313 + 0.33470 + 0.25102 = 0.80885 \end{aligned}$$

Así

$$P(X_b \geq 3) = 1 - P(X_b < 3) = 1 - P(0 \leq X_b \leq 2) = 1 - 0.80885 = 0.19115.$$

Apartado f) Ahora preguntan por un intervalo de valores de X_b, concretamente por la probabilidad de que ocurran entre 2 y 5 escapes durante el próximo año. En este caso tenemos

$$P(2 \leq X_b \leq 5) = P(X_b = 2) + P(X_b = 3) + P(X_b = 4) + P(X_b = 5).$$

De los apartados anteriores sabemos $P(X_b = 2) = 0.25102$ y $P(X_b = 3) = 0.12551$, así que falta calcular

$$\begin{aligned} P(X_b &= 4) = \frac{e^{-\lambda_b}\lambda_b^4}{4!} = \frac{e^{-1.5} \times 1.5^4}{4!} = 0.047067, \\ P(X_b &= 5) = \frac{e^{-\lambda_b}\lambda_b^5}{5!} = \frac{e^{-1.5} \times 1.5^5}{5!} = 0.01412. \end{aligned}$$

Obtenemos

$$P(2 \leq X_b \leq 5) = 0.25102 + 0.12551 + 0.047067 + 0.01412 = 0.43772.$$

Apartado g) Por último, como vuelven a cambiar el periodo de tiempo involucrado, debemos cambiar de variable aleatoria, y contar el número de accidentes por cada 3 años. Consideramos la nueva variable aleatoria $X_g \equiv$ *"número de escapes radioactivos en tres años"*. Igual que antes, obtenemos la correspondiente λ_g

$$\left\{ \begin{array}{lcl} 1 \text{ escape} & \rightarrow & 8 \text{ meses} \\ \lambda_g \text{ escapes} & \leftarrow & 3 \text{ años (36 meses)} \end{array} \right. ,$$

luego

$$\lambda_g = \frac{1 \times 36}{8} \frac{\text{escapes}}{3 \text{ años}} = 4.5 \frac{\text{escapes}}{3 \text{ años}}.$$

Asumimos ahora que X_g se distribuye como una Poisson de parámetro $\lambda_g = 4.5$, es decir $X_g \sim Poisson(\lambda_g = 4.5)$, y utilizando de nuevo (3) con el nuevo valor de λ_g tenemos que la probabilidad de sufrir entre 2 y 5 accidentes durante los próximos tres años es ahora

$$P(2 \leq X_g \leq 5) = P(X_g = 2) + P(X_g = 3) + P(X_g = 4) + P(X_g = 5),$$

donde

$$\begin{aligned} P(X_g &= 2) = \frac{e^{-\lambda_g}\lambda_g^2}{2!} = \frac{e^{-4.5} \times 4.5^2}{2!} = 0.11248, \\ P(X_g &= 3) = \frac{e^{-\lambda_g}\lambda_g^3}{3!} = \frac{e^{-4.5} \times 4.5^3}{3!} = 0.16872, \\ P(X_g &= 4) = \frac{e^{-\lambda_g}\lambda_g^4}{4!} = \frac{e^{-4.5} \times 4.5^4}{4!} = 0.18981, \\ P(X_g &= 5) = \frac{e^{-\lambda_g}\lambda_g^5}{5!} = \frac{e^{-4.5} \times 4.5^5}{5!} = 0.17083 \end{aligned}$$

y por tanto

$$P(2 \leq X_g \leq 5) = 0.11248 + 0.16872 + 0.18981 + 0.17083 = 0.64184.$$

La probabilidad es sensiblemente más alta, ya que el tiempo en el que pueden ocurrir los accidentes es tres veces mayor.

En general, las variables aleatorias que se distribuyen de acuerdo a una distribución de Poisson no tienen porqué contar únicamente sucesos en un en cierto periodo de tiempo. También pueden contar número de erratas en un las páginas de un libro, o número de células encontradas en cierto volumen de líquido. Vamos a hacer un segundo ejemplo de este tipo. Vamos a resolver uno de los problemas de refuerzo de este tema a modo de ejemplo.

Ejercicio 8 - Se sabe que el número de bacterias por cm^3 de agua en un estanque es una variable aleatoria X con distribución de Poisson de parámetro $\lambda = 0.5$ bacterias por cada cm^3. Determine las probabilidades siguientes:
a) Probabilidad de que en un cm^3 de agua del estanque no haya ninguna bacteria.
b) Probabilidad de que en un cm^3 de agua del estanque haya un máximo de 3 bacterias.
c) Probabilidad de que en 5 cm^3 de agua del estanque haya 2 bacterias o menos.
d) Probabilidad de que en 5 cm^3 de agua del estanque haya entre 2 y 5 bacterias.

Solución ejercicio 8)

Apartado a) Por ejemplo, para resolver este apartado consideramos la variable aleatoria $X_a \equiv$ *"número de bacterias en 1 cm^3 de agua del estanque"*. Averiguamos el λ_a correspondiente a este apartado, que como siempre se puede obtener planteando una regla de tres simple con el promedio que proporciona el enunciado (0.5 bacterias por 1 cm^3). Así,

$$\left\{ \begin{array}{lcl} 0.5 \text{ bacterias} & \rightarrow & \mathit{1}\ cm^3 \\ \lambda_a \text{ bacterias} & \leftarrow & \mathit{1}\ cm^3 \end{array} \right. ,$$

luego aquí no hace falta resolver la regla de tres, ya que el λ_a va a ser el mismo que se proporciona en el enunciado, es decir

$$\lambda_a = 0.5 \frac{\text{bacterias}}{cm^3}.$$

Asumimos entonces que X_a se distribuye como una Poisson de parámetro $\lambda_a = 0.5$, es decir $X_a \sim Poisson(\lambda_a = 0.5)$, y utilizando la fórmula para la función de probabilidad de la distribución de Poisson (3) tenemos que la probabilidad de que haya 0 bacterias en 1 cm^3 de agua del lago es $P(X_a = 0 \text{ bacterias})$, i.e.,

$$P(X_a = 0) = \frac{e^{-\lambda_a}\lambda_a^0}{0!} = \frac{e^{-0.5} \times 0.5^0}{0!} = e^{-0.5} = 0.60653.$$

Apartado b) Como siguen preguntando bacterias en un solo centrímetro cúbico de agua del estanque no cambiamos de variable aleatoria. Seguimos con la definida en el apartado a) y con su λ_a correspondiente. Nos preguntan en este caso

$$P(X_a \leq 3)$$

Es decir, la probabilidad de que mi X_a, que cuenta el número de bacterias en 1 cm^3, sea menor o igual que 3. Dado que el valor más bajo que puede tomar X_a es cero, esta probabilidad es equivalente a la suma

$$P(X_a \leq 3) = P(0 \leq X_a \leq 3) = P(X_a = 0) + P(X_a = 1) + P(X_a = 2) + P(X_a = 3).$$

Usando de nuevo (3) para calcular la asignación de probabilidades obtenemos

$$\begin{aligned} P(X_a &= 0) = \frac{e^{-\lambda_a}\lambda_a^0}{0!} = \frac{e^{-0.5} \times 0.5^0}{0!} = 0.60653, \\ P(X_a &= 1) = \frac{e^{-\lambda_a}\lambda_a^1}{1!} = \frac{e^{-0.5} \times 0.5^1}{1!} = 0.30327, \\ P(X_a &= 2) = \frac{e^{-\lambda_a}\lambda_a^2}{2!} = \frac{e^{-0.5} \times 0.5^2}{2!} = 0.07582, \\ P(X_a &= 3) = \frac{e^{-\lambda_a}\lambda_a^3}{3!} = \frac{e^{-0.5} \times 0.5^3}{3!} = 0.01264. \end{aligned}$$

Así pues, la probabilidad buscada es

$$P(X_a \leq 3) = 0.60653 + 0.30327 + 0.07582 + 0.01264 = 0.99826.$$

Apartado c) Como la pregunta de este apartado se refiere ahora a un volumen diferente, 5 cm^3 de agua del estanque, **tenemos que cambiar de variable aleatoria (vamos a llamarla X_c) y calcular el nuevo λ_c correspondiente**. Establecemos la nueva variable aleatoria $X_c \equiv$ *"número de bacterias en* 5 cm^3*"*. Para obtener λ_c planteamos una regla de tres simple con el promedio del enunciado. Así,

$$\left\{ \begin{array}{lcl} 0.5 \text{ bacterias} & \rightarrow & 1\ cm^3 \\ \lambda_c \text{ bacterias} & \leftarrow & 5\ cm^3 \end{array} \right. ,$$

luego

$$\lambda_c = \frac{0.5 \times 5}{1} \frac{\text{bacterias}}{5\ cm^3} = 2.5 \frac{\text{bacterias}}{5\ cm^3}.$$

La variable X_c se distribuye como una Poisson de parámetro $\lambda_c = 2.5$, es decir $X_c \sim Poisson(\lambda_c = 2.5)$, y utilizando la fórmula para la función de probabilidad correspondiente (3) tenemos que la probabilidad de encontrar 2 bacterias o menos en *5* cm^3 de agua del estanque es $P(X_c \leq 2 \text{ bacterias})$, i.e.,

$$P(X_c \leq 2) = P(0 \leq X_c \leq 2) = P(X_c = 0) + P(X_c = 1) + P(X_c = 2).$$

Debemos rehacer los cálculos de las probabilidades individuales con la λ_c correspondiente a este apartado, por lo que

$$\begin{aligned} P(X_c &= 0) = \frac{e^{-\lambda_c}\lambda_c^0}{0!} = \frac{e^{-2.5} \times 2.5^0}{0!} = 0.082085, \\ P(X_c &= 1) = \frac{e^{-\lambda_c}\lambda_c^1}{1!} = \frac{e^{-2.5} \times 2.5^1}{1!} = 0.20521, \\ P(X_c &= 2) = \frac{e^{-\lambda_c}\lambda_c^2}{2!} = \frac{e^{-2.5} \times 2.5^2}{2!} = 0.25652. \end{aligned}$$

Luego

$$P(X_c \leq 2) = 0.082085 + 0.20521 + 0.25652 = 0.54382.$$

Apartado d) Este apartado se refiere igualmente al volumen de 5 cm^3, por lo que seguimos usando la misma variable aleatoria del apartado anterior para contar número de bacterias en 5 cm^3 de agua del estanque. Ahora nos preguntan por la probabilidad de que haya entre 2 y 5 bacterias en esos 5 cm^3 estudiados, por lo que queremos calcular

$$P(2 \leq X_c \leq 5) = P(X_c = 2) + P(X_c = 3) + P(X_c = 4) + P(X_c = 5).$$

Del apartado anterior c) sabemos que $P(X_c = 2) = 0.25652$ luego queda calcular

$$\begin{aligned} P(X_c &= 3) = \frac{e^{-\lambda_c}\lambda_c^3}{3!} = \frac{e^{-2.5} \times 2.5^3}{3!} = 0.21376. \\ P(X_c &= 4) = \frac{e^{-\lambda_c}\lambda_c^4}{4!} = \frac{e^{-2.5} \times 2.5^4}{4!} = 0.1336. \\ P(X_c &= 5) = \frac{e^{-\lambda_c}\lambda_c^5}{5!} = \frac{e^{-2.5} \times 2.5^5}{5!} = 0.0668. \end{aligned}$$

Así

$$P(2 \leq X_c \leq 5) = 0.25652 + 0.21376 + 0.1336 + 0.0668 = 0.67068.$$

Ejercicio 9 - En la inspección de una hojalata producida por un proceso electrolítico continuo, se identifican en promedio $0,2$ imperfecciones por minuto.

Determine las probabilidades de identificar:

a) Una imperfección en 3 minutos.

b) Un máximo de una imperfección en 15 minutos.

Solución ejercicio 9:

El problema se resuelve utilizando una variable aleatoria que cuente el número imperfecciones, teniendo en cuenta que el promedio de ellas es de $0,2$ por minuto.

Apartado a) Consideramos la variable aleatoria $X_a \equiv$ *"número de imperfecciones en tres minutos"*, que resolviendo una regla de tres simple con el promedio que nos da el enunciado, vemos que X_a se distribuye como una Poisson de parámetro $\lambda_a = \frac{3\times 0,2}{1} = 0,6$, es decir

$$X_a \sim Poisson(\lambda_a = 0,6)$$

luego, utilizando la fórmula proporcionada en la hoja de enunciados para la distribución de Poisson

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

tenemos

$$P(X_a = 1) = \frac{e^{-\lambda_a}\lambda_a^1}{1!} = \frac{e^{-0,6}\times 0,6^1}{1!} = 0,32929.$$

$$\frac{e^{-0,6}\times 0,6^1}{1!}$$

Apartado b) Consideramos una nueva variable aleatoria $X_b \equiv$ *"número de imperfecciones en quince minutos"*, que cuenta las imperfecciones en el nuevo periodo de tiempo de quince minutos. Resolviendo de nuevo una regla de tres simple con el promedio del enunciado, vemos que X_b se distribuye como una Poisson de parámetro $\lambda_b = \frac{15\times 0,2}{1} = 3$, es decir

$$X_b \sim Poisson(\lambda_b = 3).$$

Utilizando de nuevo la función de probabilidad para la distribución de Poisson, con el nuevo parámetro $\lambda_b = 3$, la probabilidad de tener un máximo de una imperfección es $P(X_b \leq 1) = P(X = 0) + P(X = 1)$, luego

$$P(X \leq 1) = \frac{e^{-3}\times 3^0}{0!} + \frac{e^{-3}\times 3^1}{1!} = 0,049787 + 0,14936 = 0,19915.$$

Ejercicio 10 - En el sur de California se produce, en promedio, un terremoto al año de magnitud 6.1 o mayor en la escala de Richter. Terremotos de estas magnitudes son aquellos con posibilidad de provocar daños severos en zonas pobladas. Determine las probabilidades siguientes:

a) Probabilidad de que no ocurra ningún terremoto durante el próximo año.

b) Probabilidad de que ocurran entre 1 y 3 terremotos durante los próximos 15 meses.

Solución ejercicio 10.

Nota: Durante todo este problema solo nos referiremos a aquellos terremotos de magnitud 6.1 o mayor en la escala de Richter.

Apartado a) Para contar el **número de terremotos al año** definimos la variable aleatoria discreta $X \equiv$ "*número de terremotos al año en el sur de California*". Como el promedio de número de terremotos al año es de uno, tendremos $\lambda_a = 1\, terremoto/año$ y entonces $X \sim \mathcal{P}(\lambda_a = 1)$. Nos preguntan la probabilidad de que $X = 0$, luego

$$P(X = 0) = \frac{1^0 e^{-1}}{0!} = e^{-1} = 0.36788.$$

Apartado b) Como ahora se refieren a otro periodo de tiempo, **tenemos que cambiar de variable aleatoria**, para contar ahora **número de terremotos en 15 meses**. Definimos la variable aleatoria discreta $Y \equiv$ "*número de terremotos en 15 meses en el sur de California*". Calculamos la nueva λ_b haciendo una regla de tres con el promedio del enunciado

$$\left\{ \begin{array}{lcl} 1 \text{ terremoto} & \rightarrow & 12 \text{ meses} \\ \lambda_b \text{ terremotos} & \leftarrow & 15 \text{ meses} \end{array} \right. , \qquad \lambda_b = \frac{15\cdot 1}{12} = 1.25\,\frac{terremotos}{15\, meses}$$

y entonces $Y \sim \mathcal{P}(\lambda_b = 1.25)$. Nos preguntan la probabilidad de que $1 \leq Y \leq 3$ luego

$$\begin{aligned} P(1 \leq Y \leq 3) &= P\left(Y=1\right)+P\left(Y=2\right)+P\left(Y=3\right)=\frac{1.25^1e^{-1.25}}{1!}+\frac{1.25^2e^{-1.25}}{2!}+\frac{1.25^3e^{-1.25}}{3!} \\ &= 0.358\,13+0.22383+9.3263\times10^{-2}=0.67522. \end{aligned}$$

Ejercicio 11 - En una investigación sobre la presencia de zooplancton en una zona marina se observa que en esas aguas hay un promedio de seis organismos por mililitro. Determine las probabilidades siguientes:
a) Probabilidad de que haya entre 1 y 3 organismos en 1.5 mililitros de agua.
b) Probabilidad de que no haya ningún organismo en 1 mililitro de agua

Solución ejercicio 11.
Apartado a) $X \equiv$ "*número de organismos en 1.5 ml*". Con $X \sim \mathcal{P}(\lambda_a = 9)$,donde

$$\left\{ \begin{array}{lcl} 6 \text{ organismos} & \rightarrow & 1\;ml \\ \lambda_a \text{ organismos} & \leftarrow & 1.5\;ml \end{array} \right. , \qquad \lambda_a = \frac{6\cdot1.5}{1} = 9\,\frac{organismos}{ml}$$

$$\begin{aligned} P(1 \leq X \leq 3) &= P\left(X=1\right)+P\left(X=2\right)+P\left(X=3\right) \\ &= \frac{9^1e^{-9}}{1!}+\frac{9^2e^{-9}}{2!}+\frac{9^3e^{-9}}{3!} \\ &= 1.1107\times10^{-3}+4.9981\times10^{-3}+1.4994\times10^{-2}=0.021103. \end{aligned}$$

Apartado b) $Y \equiv$ "*número de organismos en 1 ml*". $\lambda_b = 6$ *organismos/ml*, $Y \sim \mathcal{P}(\lambda_a = 6)$, y nos preguntan por

$$P(Y=0)=\frac{6^0e^{-6}}{0!}=e^{-6}=0.0024788.$$

Ejercicio 12 - El Etna es el volcán más activo de Europa, para el que se vienen contabilizando un promedio de seis erupciones al año. Determine las probabilidades siguientes:
a) Probabilidad de que ocurran entre 1 y 3 erupciones durante los próximos 18 meses.
b) Probabilidad de que no ocurra ninguna erupción durante el próximo año.

Solución ejercicio 12:
Apartado a) Para contar el **número de erupciones durante los próximos 18 meses** definimos la variable aleatoria discreta $X \equiv$ "*número de erupciones del Etna los próximos 18 meses*". Como el promedio de número de erupciones al año (cada 12 meses) es de seis, tendremos

$$\left\{ \begin{array}{lcl} 6 \text{ erupciónes} & \rightarrow & 12 \text{ meses} \\ \lambda_a \text{ erupciones} & \leftarrow & 18 \text{ meses} \end{array} \right. , \qquad \lambda_b = \frac{18\cdot6}{12} = 9\,\frac{erupciones}{18\,meses}$$

$\lambda_a = 9\,erupciones/18\,meses$ y entonces $X \sim \mathcal{P}(\lambda_a = 9)$. Nos preguntan la probabilidad de que $1 \leq X \leq 3$, luego

$$\begin{aligned} P(1 \leq X \leq 3) &= P\left(X=1\right)+P\left(X=2\right)+P\left(X=3\right)=\frac{9^1e^{-9}}{1!}+\frac{9^2e^{-9}}{2!}+\frac{9^3e^{-9}}{3!} \\ &= 1.1107\times10^{-3}+4.9981\times10^{-3}+1.4994\times10^{-2}=0.021103. \end{aligned}$$

Apartado b) Como ahora se refieren a otro periodo de tiempo, **tenemos que cambiar de variable aleatoria**, para contar ahora **número de erupciones del Etna en un año**. Definimos la variable aleatoria discreta $Y \equiv$ "*número de erupciones del Etna en un año*". No hace falta calcular la nueva λ_b porque es exactamente el promedio que nos dan en el enunciado, de seis erupciones al año, luego $Y \sim \mathcal{P}(\lambda_a = 6)$, y nos preguntan por

$$P(Y=0)=\frac{6^0e^{-6}}{0!}=e^{-6}=0.0024788,$$

que es una probabilidad muy baja. Se puede dar casi por seguro que habrá erupciones durante el próximo año.