1. Enunciado

Sea X una variable aleatoria con distribución de Poisson(λ) de la que se tomó una muestra aleatoria simple de tamaño $n = 10$ para contrastar las hipótesis $H_0\colon \lambda = \lambda_0 = 2$ y $H_1\colon \lambda = \lambda_1 = 4$ a nivel de significación $\alpha = 0{,}05$.

1. Usa el lema de Neyman-Pearson para construir un contraste para esas hipótesis. Compara su tamaño y el nivel de significación.
2. Calcula el p-valor del contraste para la muestra $(2, 1, 4, 1, 1, 3, 1, 0, 1, 2)$.
3. Construye un nuevo contraste cuyo tamaño coincida con el nivel de significación y calcula su potencia bajo H_1.
4. Simula por Montecarlo el tamaño de los dos contrastes.

2. Resolución

2.1. Contraste no aleatorizado de Neyman-Pearson

Las funciones de verosimilitud son:

- bajo H_0: $L_0(x_1, \ldots, x_n) = e^{-n\lambda_0} \frac{\lambda_0^{\sum x_i}}{\prod x_i!}$
- bajo H_1: $L_1(x_1, \ldots, x_n) = e^{-n\lambda_1} \frac{\lambda_1^{\sum x_i}}{\prod x_i!}$

La razón de verosimilitudes es:

$$\frac{L_1(x_1, \ldots, x_n)}{L_0(x_1, \ldots, x_n)} = \frac{e^{-n\lambda_1} \frac{\lambda_1^{\sum x_i}}{\prod x_i!}}{e^{-n\lambda_0} \frac{\lambda_0^{\sum x_i}}{\prod x_i!}} = e^{-n(\lambda_1-\lambda_0)} \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum x_i}$$

La región crítica del contraste es:

$$\begin{aligned} \left\{\vec{x} \;\middle|\; \frac{L_1(x_1, \ldots, x_n)}{L_0(x_1, \ldots, x_n)} > k\right\} &= \left\{\vec{x} \;\middle|\; e^{-n(\lambda_1-\lambda_0)} \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum x_i} > k\right\} \\ &= \left\{\vec{x} \;\middle|\; \sum x_i > k_1\right\} \subset C_0 \end{aligned}$$

pues $4 = \lambda_1 > \lambda_0 = 2$.

El valor de λ_1 se calcula teniendo en cuenta el nivel de significación:

$$\Pr(\text{R.C.} \mid H_0) = \Pr\left(\sum x_i > k_1 \mid H_0\right) = \Pr[\text{Poisson}(n\lambda_0) > k_1]$$

```
> (k1 <- qpois (1-alfa, n*landa0))
```

```
[1] 28
```

Teniendo en cuenta el resultado, se pueden plantear dos regiones críticas:

$$\text{R.C.1} = \left[\sum X_i > 27\right] \qquad \text{R.C.2} = \left[\sum X_i > 28\right]$$

cuyos tamaños son

- $\Pr\left[\text{R.C.1} \mid H_0\right] = \Pr\left[\sum X_i > 27 \mid H_0\right] = \Pr\left[\text{Poisson}(\lambda_0) > 27\right]$

  ```
  >       1 - ppois (k1-1, n*landa0)
  ```

  ```
  [1] 0.05248071
  ```

  que sobrepasa el nivel de significación.

- $\Pr\left[\text{R.C.2} \mid H_0\right] = \Pr\left[\sum X_i > 28 \mid H_0\right] = \Pr\left[\text{Poisson}(\lambda_0) > 28\right]$

  ```
  >     1 - ppois (k1, n*landa0)
  ```

  ```
  [1] 0.03433352
  ```

  que cumple la condición asociada con α.

Por tanto, el contraste elegido tiene como región crítica

$$\left[\sum X_i > 28\right] = \left[\sum X_i \geq 29\right] = \left[\sum X_i \geq 28{,}5\right] = \left[\sum X_i > 28{,}5\right]$$

2.2. Potencia

Es el contraste más potente para una significación de 0,03433352 y su potencia vale

```
>     1 - ppois (k1, n*landa1)
```

```
[1] 0.9706204
```

2.3. P-valor

```
> X
```

```
 [1] 2 1 4 1 1 3 1 0 1 2
```

```
> sum(X)
```

```
[1] 16
```

```
> 1 - ppois (sum(X)-1, n*landa0) # Pr[Poisson(n landa0) >= sum(X)]
```

```
[1] 0.8434869
```

2.4. Contraste aleatorizado de Neyman-Pearson

El contraste anterior tiene un tamaño = 0,03433352 estrictamente menor que el nivel de significación y lo que se pretende conseguir con un contraste aleatorizado es que ambos coincidan.

Puesto que $\Pr[\text{R.C.} \mid H_0] < \alpha$ es necesario aumentar, de alguna manera, la región crítica con muestras que sean mucho menos verosímiles bajo H_0 que bajo H_1. Las candidatas naturales para ser incluidas en la nueva región crítica son las que están en la frontera, es decir, aquéllas que cumplen $\sum x_i = 28$.

El problema es que, entonces, la probabilidad de la nueva región crítica sería mayor que α. Por eso se usa un procedimiento aleatorizado que da lugar a la siguiente región crítica:

$$\text{R.C.a.} = \left[\sum X_i > 28\right] \bigcup \left[\sum X_i = 28 \cap U < k_2\right]$$

donde U es una variable Uniforme$(0,1)$ independiente de las X_i.

$$\begin{aligned}
\Pr[\text{R.C.a.} \mid H_0] &= \Pr_{H_0}\left(\left[\sum X_i > 28\right] \bigcup \left[\sum X_i = 28 \cap U < k_2\right]\right) \\
&= \Pr_{H_0}\left[\sum X_i > 28\right] + \Pr_{H_0}\left[\sum X_i = 28 \cap U < k_2\right] \\
&= \Pr_{H_0}\left[\sum X_i > 28\right] + \Pr_{H_0}\left[\sum X_i = 28\right] \Pr[U < k_2] \\
\Rightarrow \quad k2 = \Pr[U < k_2] &= \frac{\Pr[\text{R.C.a.} \mid H_0] - \Pr[\sum X_i > 28 \mid H_0]}{\Pr[\sum X_i = 28 \mid H_0]} \\
&= \frac{\alpha - \text{tamaño}}{\Pr[\text{Poisson}(\lambda_0) = 28]}
\end{aligned}$$

```
> tama <- 1 - ppois (k1, n*landa0)
> (k2 <- (alfa - tama) / dpois (k1, n*landa0))
```

```
[1] 0.8633004
```

Las potencias de ambos contrastes serían:

- determinístico (calculada antes)

  ```
  > 1 - ppois (k1, n*landa1)
  ```

  ```
  [1] 0.9706204
  ```

- aleatorizado

  $$\Pr[\text{R.C.a} \mid \lambda_1] = \Pr_{H_1}\left[\sum X_i > 28\right] + \Pr_{H_1}\left[\sum X_i = 28\right] k_2$$

  ```
  > 1-ppois(k1,n*landa1) + dpois(k1,n*landa1)*k2
  ```

  ```
  [1] 0.9792885
  ```

2.5. Simulación

```
> rechazos <- replicate (1e5,
+ {
+     X <- rpois (n, landa0)
+     suma <- sum(X)
+     c (deter = suma > k1,
+        aleat = (suma > k1) + (suma == k1) * (runif(1) < k2))
+ })
> apply (rechazos, 1, mean)   # tamaños simulados

  deter   aleat
0.03430 0.04994

> c (deter=tama, aleat=alfa)  # tamaños teóricos

     deter      aleat
0.03433352 0.05000000
```