Se quiere estimar el número de fábricas capaces de producir armas bacteriológicas en un determinado país de 36.000 Km cuadrados de extensión. Para ello se envió un avión espía U2 sobre el mencionado país enemigo que realizó 250 fotografías aéreas correspondientes a 250 parcelas elegidas al azar, cada una de ellas de dimensión 30 Km cuadrados, detectándose una planta química entre las fotografías realizadas por el mencionado avión espía. ¿Qué estimación daría para el número de plantas químicas del país?

El problema puede formularse de varias maneras según la modelización que se haga del mismo. Parece razonable dividir el país en parcelas del tamaño tomado por las fotografías; es decir, en 1200 parcelas, cada una de dimensión 30 Km cuadrados. Estamos, en ese caso, ante una *población* de tamaño 1200 de la que tomamos una muestra aleatoria de tamaño $n = 250$.

Sin utilizar las herramientas suministradas por la *Teoría de Muestras para Poblaciones Finitas*, la cual permite obtener conclusiones para muestras tomadas sin reemplazamiento, admitiremos aquí que las 250 parcelas fotografiadas por el avión espía fueron elegidas al azar con reemplazamiento no resultando elegida ninguna de ellas en más de una ocasión.

Supondremos además que el país enemigo en cuestión repartió las plantas químicas por su territorio al azar confiando que la aleatoriedad no diera pistas a los servicios secretos enemigos y que, en cada parcela de 30 Km cuadrados, o bien hay una planta o bien no hay ninguna planta química. En estas condiciones, el número de plantas químicas del país es una variable aleatoria que sigue una distribución binomial $B(1200, p)$ siendo p la probabilidad de que en cada parcela haya planta química, probabilidad que supondremos constante de parcela a parcela. Es decir, suponemos que la colocación de una planta química en una parcela es un experimento de Bernoulli en donde se dan dos situaciones posibles: *éxito*, que haya planta química, o *fracaso* que no la haya, siendo la probabilidad de *éxito* p y supuesto que realizamos 1200 pruebas de ese tipo independientemente, manteniendo constante la probabilidad de *éxito*.

En esta situación, el parámetro a estimar puede ser la media de la binomial, $n \cdot p = 1200 \cdot p$, que representaría el número medio de plantas químicas en el país en cuestión después de haber *jugado* las 1200 pruebas de Bernoulli.

Otra posibilidad es la de determinar el valor más probable que puede tomar X; es decir, calcular $P\{X = x\}$ para los diversos x y elegir como estimación del número de plantas químicas, aquel valor x para el que la probabilidad anterior sea mayor, es decir, la moda de la distribución. En ambos casos necesito determinar un estimador de p, parámetro de una distribución de Bernoulli $X_i \rightsquigarrow B(1,p)$ siendo $X = \sum_{i=1}^{1200} X_i \rightsquigarrow B(1200,p)$.

Para ello utilizaremos el método de la máxima verosimilitud. Como la función de masa de la Bernoulli $B(1,p)$ es (la binomial $B(1,p)$ recibe el nombre de distribución de Bernoulli)

$$p(x) = p^x\,(1-p)^{1-x} \qquad\qquad x = 0,1$$

la función de verosimilitud de la muestra será

$$L(p) = \prod_{i=1}^{1200} p^{x_i}\,(1-p)^{1-x_i}$$

de logaritmo

$$\log L(p) = \sum_{i=1}^{1200} \left[x_i \log p + (1-x_i)\log(1-p)\right].$$

Su derivada igualada a cero —ecuación de verosimilitud— será

$$\frac{d}{dp}\log L(p) = \sum_{i=1}^{1200}\left[\frac{x_i}{p} - (1-x_i)\frac{1}{1-p}\right] = 0.$$

es decir,

$$\frac{\sum^{1200} x_i}{p} - \frac{1200 - \sum^{1200} x_i}{1-p} = 0$$

o bien,

$$\hat{p} = \frac{\sum^{1200} x_i}{1200}$$

es decir, la proporción muestral. Además conocemos su distribución en el muestreo, ya que $1200\hat{p} = \sum_{i=1}^{1200} X_i$ es una distribución binomial $B(1200,p)$. Este hecho permite determinar sus propiedades, como por ejemplo que es insesgado; es decir, que sea cual sea el valor de p,

$$E[\hat{p}] = \frac{1}{1200}E[\sum_{i=1}^{1200} X_i] = \frac{1}{1200}\sum_{i=1}^{1200} E[X_i] = \frac{1}{1200}\sum_{i=1}^{1200} 1\cdot p = \frac{1}{1200}1200\cdot p = p.$$

Por tanto, una estimación insesgada de p será $\hat{p} = 1/250 = 0'004$ y del número de plantas químicas del país, $1200 \cdot \hat{p} = 1200 \cdot 0'004 = 4'8$, estimación que presenta el inconveniente de no ser un número entero.

Vamos a realizar la otra estimación del número más probable de plantas químicas. Para ello debemos determinar las probabilidades $P\{X = x\}$ para diversos x siendo $X \rightsquigarrow B(1200, 0'004)$. En esta situación de ser el primer parámetro de la binomial, 1200, muy grande en relación al producto de los dos parámetros, $1200 \cdot 0'004 = 4'8$, y este producto muy grande en relación al segundo parámetro, $0'004$, puede aproximarse la distribución binomial por una distribución de Poisson (CB-sección 4.4.2) de parámetro $1200 \cdot 0'004 = 4'8$. Por tanto, será $P\{X = x\} = P\{Y = x\}$ con $Y \rightsquigarrow \mathcal{P}(4'8)$. La función de masa de la distribución anterior será

$$p(x) = \frac{e^{-4'8} \cdot 4'8^x}{x!} = \frac{0'00823 \cdot 4'8^x}{x!}.$$

Esta función de x crece hasta un valor máximo y luego empieza a decrecer, al igual que ocurre con la función de masa de la distribución de Poisson para otros valores del parámetro que sí están tabulados en la tabla 2. Como es

$$\begin{aligned}
p(0) &= \frac{0'00823 \cdot 4'8^0}{0!} = 0'0082. \\
p(1) &= \frac{0'00823 \cdot 4'8^1}{1!} = 0'0395. \\
p(2) &= \frac{0'00823 \cdot 4'8^2}{2!} = 0'0948. \\
p(3) &= \frac{0'00823 \cdot 4'8^3}{3!} = 0'1517. \\
p(4) &= \frac{0'00823 \cdot 4'8^4}{4!} = 0'1820. \\
p(5) &= \frac{0'00823 \cdot 4'8^5}{5!} = 0'1748.
\end{aligned}$$

concluimos que el valor más probable de plantas químicas del país es 4. Como los valores próximos a la moda, 3 y 5, tienen probabilidades muy similares a la del valor 4, esta medida de posición no es tan representativa de la distribución de probabilidad como la media. Por tanto, consideramos a la estimación de la media más representativa y, como no es un número entero, concluimos con una estimación del número de plantas químicas igual a 5.