SOLUCIONES PROBLEMAS ESTADÍSTICA

1 Supongamos que medimos $k = 100$ eventos, de probabilidad muy pequeña en un intervalo de tiempo dado. ¿Cuál será la desviación estándar σ? Estamos midiendo una variable aleatoria, y el valor que obtengamos en cada medida estará en torno al "valor auténtico" μ. Podemos decir con una certeza del 90 % que el valor "verdadero" μ está entre dos límites. Esto es, el 90 % del área de la distribución de probabilidad subyacente está entre esos dos límites. ¿Cuáles son tales límites? Suponga que, en lugar de $k = 100$ eventos, hubiéramos medido sólo $k = 1$. Si el valor "verdadero" fuera $k = 1$, ¿cuál es la probabilidad de medir un evento o menos de uno? ¿Cuál es la probabilidad de medir más de un evento? ¿Cuáles son los límites de certeza que podemos asignar a μ?

Experimentos aleatorios con 2 resultados posibles efectuados en un intervalo de tiempo dado. Resultados posibles: medición de evento o no medición de evento.

Ejemplo: tenemos un detector de ondas gravitacionales y lo dejamos funcionando durante 1 hora. El detector registrará eventos pero algunos de ellos no serán "detecciones reales" sino fluctuaciones espúreas del detector.

Las condiciones descritas implican que debemos adoptar la distribución de Poisson como distribución de población:

$$P(x|\mu) = \frac{\mu^x}{x!} e^{-\mu}$$

μ = nº de medidas reales de la magnitud que estamos midiendo en el intervalo de tiempo considerado.

Recordemos que en la distribución de Poisson la variable aleatoria x representa el número de eventos registrados. El valor de μ es desconocido y habrá que estimarlo a través de un estimador. Definimos el siguiente estimador:

$\hat{\mu} = x$ (lo que esto quiere decir es que suponemos que todos los eventos registrados son reales)

Distribución muestral $P(\hat{\mu}|\mu) = \dfrac{\mu^{\hat{\mu}}}{\hat{\mu}!} e^{-\mu}$, $\hat{\mu} = 0, 1, 2, \ldots$

Se tiene que $E(\hat{\mu}) = \mu \Rightarrow \hat{\mu}$ es un estimador no sesgado.

La desviación típica de la distribución de Poisson es $\sigma = \sqrt{\mu}$ así que una estimación de μ nos da automáticamente una estimación de σ, que llamaremos $\hat{\sigma}$. Así que $\hat{\sigma} = \sqrt{\hat{\mu}}$.

En el ejercicio $\hat{\mu}_{obs} = 100 \Rightarrow \hat{\sigma}_{obs} = \sqrt{100} = 10$

* Cálculo del intervalo de confianza: aproximamos la distribución muestral por una normal:

$$P(\hat{\mu}|\mu) = \frac{\mu^{\hat{\mu}}}{\hat{\mu}!} e^{-\mu} \Rightarrow \text{Si } \mu \gg 1 \Rightarrow P(\hat{\mu}|\mu) \approx \frac{1}{\sqrt{2\pi\mu}} \exp\left[-\frac{(\hat{\mu}-\mu)^2}{2\mu}\right]$$

$$\hat{\mu} \sim N(\mu, \mu^{\frac{1}{2}})$$

$$\begin{cases} \displaystyle\int_{-\infty}^{\hat{\mu}_{obs}} d\hat{\mu} \frac{1}{\sqrt{2\pi\mu_+}} \exp\left[-\frac{(\hat{\mu}-\mu_+)^2}{2\mu_+}\right] = \alpha \Rightarrow \Phi\left[\frac{\hat{\mu}_{obs}-\mu_+}{\mu_+^{\frac{1}{2}}}\right] = \alpha \\ \\ \displaystyle\int_{\hat{\mu}_{obs}}^{\infty} d\hat{\mu} \frac{1}{\sqrt{2\pi\mu_-}} \exp\left[-\frac{(\hat{\mu}-\mu_-)^2}{2\mu_-}\right] = \beta \Rightarrow 1-\Phi\left[\frac{\hat{\mu}_{obs}-\mu_-}{\mu_-^{\frac{1}{2}}}\right] = \beta \end{cases}$$

usar $\Phi^{-1}(z) = -\Phi^{-1}(1-z)$

$$\Rightarrow \hat{\mu}_{obs} - \mu_+ = \mu_+^{\frac{1}{2}}\Phi^{-1}(\alpha) \Rightarrow \mu_+ + \mu_+^{\frac{1}{2}}\Phi^{-1}(\alpha) - \hat{\mu}_{obs} = 0$$

$$\Phi^{-1}(1-\beta) = \frac{\hat{\mu}_{obs} - \mu_-}{\mu_-^{\frac{1}{2}}} \Rightarrow \mu_-^{\frac{1}{2}}\Phi^{-1}(1-\beta) = \hat{\mu}_{obs} - \mu_- \Rightarrow \mu_- + \mu_-^{\frac{1}{2}}\Phi^{-1}(1-\beta) - \hat{\mu}_{obs} = 0$$

$$\begin{cases} \mu_+ - \mu_+^{\frac{1}{2}}\Phi^{-1}(1-\alpha) - \hat{\mu}_{obs} = 0 \Rightarrow \mu_+^{\frac{1}{2}} = \dfrac{\Phi^{-1}(1-\alpha) + \left(\Phi^{-1}(1-\alpha)^2 + 4\hat{\mu}_{obs}\right)^{\frac{1}{2}}}{2} \\ \mu_- + \mu_-^{\frac{1}{2}}\Phi^{-1}(1-\beta) - \hat{\mu}_{obs} = 0 \Rightarrow \mu_-^{\frac{1}{2}} = \dfrac{-\Phi^{-1}(1-\beta) + \left(\Phi^{-1}(1-\beta)^2 + 4\hat{\mu}_{obs}\right)^{\frac{1}{2}}}{2} \end{cases}$$

El intervalo de confianza vendría dado por $[\mu_-, \mu_+]$

Aplicación numérica:

$\alpha = \beta \Rightarrow$ nivel de confianza $= 1 - 2\alpha = 0.9 \Rightarrow 0.1 = 2\alpha \Rightarrow \alpha = 0.05$ $1-\alpha = 0.95$

$\Phi^{-1}(0.95) \sim 1.6$, $\hat{\mu}_{obs} = 100 \Rightarrow \mu_+ = 117$

$\mu_- = 85$

Si **SUPONEMOS** que las 100 detecciones que hemos obtenido son todas "detecciones reales" entonces podemos afirmar con un nivel de confianza del 90% que en experimentos futuros, el número inferior de detecciones reales será de 85 y el número superior de detecciones reales será de 117.

Si el valor verdadero es $k=1$ entonces eso es suponer que conocemos μ y que $\mu=1$ $\Rightarrow$ no hay que estimar nada.

Distribución de población $\Rightarrow$ $P(x|\mu) = \frac{\mu^x}{x!} e^{-\mu}$ $\Rightarrow$ $\mu = 1$

Probabilidad de medir un evento o menos de uno $\Rightarrow$ $P_r(x \leq 1|\mu) = e^{-1} + e^{-1} = \frac{2}{e} \approx 0.736$

Probabilidad de medir más de un evento $\Rightarrow$ $P_r(x > 1|\mu) = 1 - P_r(x \leq 1|\mu) = 1 - \frac{2}{e} \approx 0.264$

Intervalo de confianza: dado que por hipótesis μ es conocido entonces el intervalo de confianza para el número de detecciones se calcularía utilizando directamente la distribución de población:

$$\frac{90}{100} = P_r(x_- \leq x \leq x_+) = \sum_{x=x_-}^{x=x_+} P(x|\mu) = \sum_{x=x_-}^{x=x_+} \frac{\mu^x}{x!} e^{-\mu} = \frac{1}{e} \sum_{x=x_-}^{x=x_+} \frac{1}{x!} \Rightarrow$$

Tenemos que encontrar valores de x_- y x_+ que cumplan la ecuación anterior. Como x_- y x_+ son números naturales nunca conseguiremos valores que resuelvan la ecuación de manera exacta y por tanto tenemos que buscar valores que la resuelvan de forma aproximada. Para hacerlo vamos probando valores de x_- y x_+ hasta hallar una solución aproximada:

Tenemos que estudiar la siguiente ecuación:

$$e \times 0.9 \sim 2.246 = \underbrace{\frac{1}{(x_-)!} + \frac{1}{(x_-+1)!} + \dots \frac{1}{(x_-+h)!}}_{} \quad \text{donde } x_- + h = x_+$$

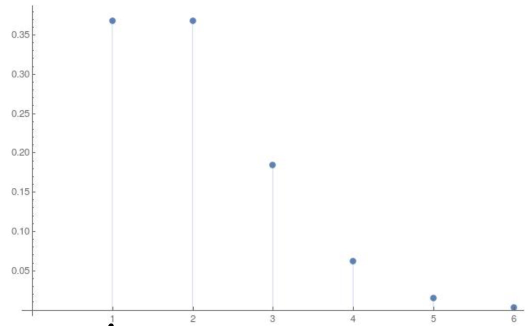
Vamos probando valores en esta ecuación

Si $x_- = 0 \Rightarrow$

$h=1 \Rightarrow 2.246 > 1 + 1$ (no vale)

$h=2 \Rightarrow 2.246 < 1 + 1 + \frac{1}{2}$ (sí vale)

tomamos $h=2$ y quedaría $x_- = 0$, $x_+ = 2$



Función de probabilidad $P(x|\mu)$, $\mu = 1$

Si $x_- = 1 \Rightarrow$

$h=1$

$2.246 > 1 + \frac{1}{2}$

$h=2$

$2.246 > 1 + \frac{1}{2} + \frac{1}{3!}$

$\vdots$

$\Rightarrow$ ningún valor de h nos vale

Para cualquier valor de h $\; 2.246 > 1 + \frac{1}{2} + \dots + \frac{1}{(1+h)!}$ porque

$$1 + \frac{1}{2} + \dots + \frac{1}{(1+h)!} < \sum_{h=0}^{\infty} \frac{1}{(1+h)!} = e - 1 = 1.718 < 2.246$$

Al final haciendo el mismo razonamiento se prueba que ningún valor de $x_- \geq 1$ nos va a valer. Esto se ve intuitivamente mirando al gráfico, que es lo que yo hice en clase. Por tanto el intervalo de confianza es $[0,2]$

Si **SABEMOS** que la única detección que hemos obtenido es una detección de verdad entonces podemos afirmar con un nivel de confianza del 90% que en experimentos futuros, el número inferior de detecciones reales será 0 y el número superior de detecciones reales será de 2.

Se puede también resolver el problema imaginando que NO SABEMOS si la única detección es real o no. Esto es equivalente a suponer que no conocemos el parámetro λ de la distribución de Poisson y por tanto habrá que estimarlo.

Distribución muestral $\Rightarrow$ $P(\hat{\lambda}|\lambda) = \frac{\lambda^{\hat{\lambda}}}{\hat{\lambda}!} e^{-\lambda}$ $\Rightarrow$ $\begin{cases} \sum_{\hat{\lambda}=0}^{\hat{\lambda}=\hat{\lambda}_{obs}} \frac{\lambda_+^{\hat{\lambda}}}{\hat{\lambda}!} e^{-\lambda_+} = \alpha \\ \sum_{\hat{\lambda}=\hat{\lambda}_{obs}}^{\infty} \frac{\lambda_-^{\hat{\lambda}}}{\hat{\lambda}!} e^{-\lambda_-} = \beta \end{cases}$

$\hat{\lambda}_{obs} = 1 \Rightarrow$

$$\sum_{\hat{\lambda}=0}^{1} \frac{\lambda_+^{\hat{\lambda}}}{\hat{\lambda}!} e^{-\lambda_+} = e^{-\lambda_+}(1+\lambda_+) = \alpha \quad || \quad \lambda_+ - \alpha e^{\lambda_+} + 1 = 0$$

$$\sum_{\hat{\lambda}=1}^{\infty} \frac{\lambda_-^{\hat{\lambda}}}{\hat{\lambda}!} e^{-\lambda_-} = e^{-\lambda_-}\left(\sum_{\hat{\lambda}=0}^{\infty} \frac{\lambda_-^{\hat{\lambda}}}{\hat{\lambda}!} - 1\right) = e^{-\lambda_-}(e^{\lambda_-} - 1) = \beta$$

$$= 1 - e^{-\lambda_-} = \beta \Rightarrow \log[1-\beta] = -\lambda_- \Rightarrow \lambda_- = -\log[1-\beta]$$

Intervalo central de confianza $\Rightarrow \alpha = \beta$

$1-\alpha-\beta = 0.9 \quad 1-2\alpha = 0.9 \quad 0.1 = 2\alpha \Rightarrow \alpha = 0.05 = \beta$

$$\left.\begin{array}{l} \lambda_+ - 0.05\, e^{\lambda_+} + 1 = 0 \Rightarrow \lambda_+ \sim 4.74 \\ \lambda_- = -\log[1-0.05] = -\log[0.95] \sim 0.01 \end{array}\right\} \Rightarrow [0, 4]$$

$\boxed{2}$ En un conjunto de experimentos idénticos, se ha atrapado un número pequeño de átomos en una trampa. En cada ocasión el número de átomos atrapados obedece a una distribución de Poisson. Es evidente de los resultados que el valor promedio es 16. ¿Cuál es la probabilidad de que el número de átomos atrapados esté entre 12 y 20? ¿Cuál sería esa probabilidad si la distribución fuera gaussiana con media y varianza ambas de valor 16?

Es evidente de los resultados $\Rightarrow$ SABEMOS el valor medio así que no hay que estimar nada y podemos usar directamente la distribución de población.

i) Distribución de población = Distribución de Poisson $P(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, $\lambda = 16$

$x \equiv$ nº de átomos atrapados

$$Pr(12 \le x \le 20) = e^{-\lambda} \sum_{x=12}^{20} \frac{\lambda^x}{x!} = 0.675046$$

ii) Distribución normal $N(\mu, \sigma) \Rightarrow P(x|\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$

$\mu = \sigma = 16$

$$Pr(12 \le x \le 20) = \int_{12}^{20} p(x)\, dx \approx 0.682689$$

$\boxed{3}$ Según los cálculos de una empresa productora de bombillas, hay una probabilidad del 0,5 % de que una bombilla sea defectuosa. Supongamos que adquirimos una muestra aleatoria de 200 bombillas. Calcule la probabilidad de que haya 0, 1 o 2 bombillas defectuosas en la citada muestra. Utilice la distribución binomial, y compare lo que resultaría de la distribución de Poisson. Sin hacer cálculos, ¿podría haber sugerido de partida que el uso de la distribución de Poisson iba a ser una buena aproximación?

De nuevo sabemos los parametros de las distribuciones de probabilidad de forma exacta:

Distribución binomial $P(x|n,\lambda) = \binom{n}{x} r^x (1-r)^{n-x}$ $n = 200$, $r = 0.005$

$$Pr(0 \le x \le 2) = 0.920161$$

Distribución de Poisson $P(x;\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$ $\lambda = 200 \times 0.005 = 2 \times 0.5 = 1$

$P_r(0 \leq x \leq 2) =$ 0.919699

Podríamos haber sugerido el uso de la distribución de Poisson, porque aproxima a la binomial si el número de experimentos aleatorios es grande y la probabilidad de éxito tiende a cero

$$\mathrm{Pois}(\lambda) = \lim_{n\to\infty} \mathrm{Bin}(n, p) \quad \text{si } p \to 0 \quad (\lambda = np)$$

4 Compruebe que el estadístico

$$\hat{\mu} = \sum_{i=1}^{n} \left(\frac{n+1}{n^2}\right) x_i$$

es un estimador consistente de la media de la distribución de la población subyacente que supondremos normal. Asimismo, calcule el sesgo del estimador y su varianza.

* Se puede hacer sin suponer que la distribución de población es normal, basta suponer que trabajamos con distribuciones de población que cumplen las condiciones estipuladas en clase:

- Consistencia: $\lim_{n\to\infty} \hat{\mu} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mu$

- Sesgo $\langle\hat{\mu}\rangle = \left(\frac{n+1}{n^2}\right) \sum_{i=1}^{n} \langle x_i \rangle = \left(\frac{n+1}{n^2}\right) n \mu \Rightarrow b(\mu) = \langle\hat{\mu}\rangle - \mu$

$$b(\mu) = \left(\frac{n+1}{n^2}\right) n\mu - \mu = \mu\left[\frac{n+1}{n} - 1\right] = \frac{\mu}{n}$$

- Varianza $\sigma^2_{\hat{\mu}} = \langle(\hat{\mu} - \langle\hat{\mu}\rangle)^2\rangle = \langle\hat{\mu}^2 + \langle\hat{\mu}\rangle^2 - 2\hat{\mu}\langle\hat{\mu}\rangle\rangle =$

$= \langle\hat{\mu}^2\rangle - \langle\hat{\mu}\rangle^2$ Falta por calcular $\langle\hat{\mu}^2\rangle$

$$\hat{\mu} = \left(\frac{n+1}{n^2}\right)\sum_{i=1}^{n} x_i \Rightarrow \hat{\mu}^2 = \frac{(n+1)^2}{n^4}\sum_{i,j}^{n} x_i x_j \Rightarrow \langle \hat{\mu}^2 \rangle = \frac{(n+1)^2}{n^4}\sum_{i,j}^{n} \langle x_i x_j \rangle =$$

$$= \frac{(n+1)^2}{n^4}\left[\sum_{\substack{i,j \\ (\neq)}}^{n} \langle x_i x_j \rangle + \sum_{i=1}^{n} \langle x_i^2 \rangle\right] = \frac{(n+1)^2}{n^4}\left[\sum_{\substack{i,j \\ (\neq)}}^{n} \underbrace{\langle x_i \rangle}_{\mu} \underbrace{\langle x_j \rangle}_{\mu} + \sum_{i=1}^{n} \langle x_i^2 \rangle\right] =$$

$$= \frac{(n+1)^2}{n^4}\left[n(n-1)\mu^2 + \sum_{i=1}^{n}\left(\langle x_i^2 \rangle - \langle x_i \rangle^2 + \langle x_i \rangle^2\right)\right] =$$

$$= \frac{(n+1)^2}{n^4}\left[n(n-1)\mu^2 + n\mu^2 + n\sigma^2\right] = \frac{(n+1)^2}{n^4}\left[n^2\mu^2 + n\sigma^2\right] =$$

$$= \frac{(n+1)^2}{n^4} n\left[\sigma^2 + n\mu^2\right] = \frac{(n+1)^2}{n^3}\left[\sigma^2 + n\mu^2\right]$$

Por tanto $\sigma^2_{\hat{\mu}} = \frac{(n+1)^2}{n^3}\left[\sigma^2 + n\mu^2\right] - \left(\frac{n+1}{n^2}\right)^2 n^2\mu^2 =$

$$= \frac{(n+1)^2}{n^3}\sigma^2$$

8 En un experimento se realizan las siguientes medidas de una magnitud:

101, 100, 97, 98, 103, 107, 98, 102, 99, 101, 100, 97.

a) Estime la media y la varianza de la distribución de la población subyacente.
b) Utilizando el hecho de que el número de datos recogidos es grande, calcule el intervalo central de confianza del 97 % para la media.

a) Estimación de la media $\hat{\mu}_{obs} = \frac{1}{12}(101+100+97+98+103+107+98+102+99+101+100+97)$

$\hat{\mu}_{obs} = \frac{401}{4} \sim 100$

- Estimación de la varianza: para estimar la varianza utilizamos el siguiente estimador:

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{\mu}_{obs})^2 \Rightarrow$$ estimador sin sesgo de la varianza

Por tanto

redondeamos para tener el mismo número de cifras significativas que en los datos

$$\hat{\sigma}^2 = \frac{1}{N}\sigma^2 \Rightarrow \sigma^2 = \frac{361}{44} \approx 8$$

cuando se sustituye $\hat{\mu}_{obs} = \frac{401}{4}$ en la fórmula de σ^2 sale este número.

b) La fórmula que hay que aplicar es la siguiente:

$$\mu_+ = \hat{\mu}_{obs} + \hat{\sigma}\,\Phi^{-1}[1-\alpha]$$

$$\mu_- = \hat{\mu}_{obs} - \hat{\sigma}\,\Phi^{-1}[1-\alpha]$$

$1-2\alpha = 0.97 \Rightarrow 2\alpha = 0.03 \Rightarrow \alpha = 0.015$

$1-\alpha = 0.985 \Rightarrow \Phi^{-1}(0.985) = 2.17$

Por tanto $\mu_+ = 100 + 0.8 \times 2.17 = 102$

$\mu_- = 100 - 0.8 \times 2.17 = 98$

$\boxed{6}$ Un medicamento bien establecido cura un 80 % de los casos tratados. Se propone un tratamiento alternativo y se desea evaluar su eficacia. En una prueba clínica con 100 pacientes 85 de los casos presentan curación. ¿Puede excluir al 95 % de confianza que este nuevo medicamento sea más eficaz que el primer tratamiento? Sugerencia: obtenga la probabilidad de distribución de que al aplicar el tratamiento a n pacientes k se curen, y considere que en este caso $n = 100 >> 1$.

Distribución de probabilidad de que al aplicar el tratamiento a n pacientes k se curen: $B(n, \frac{k}{n}) = B(n, p)$, $p = \frac{k}{n}$

Aplicamos la aproximación normal de la binomial:

$$B(n,p) \to N(np, (np(1-p))^{1/2}) \quad (n>>1)$$

En nuestro caso la aproximación da $N(k, (k(1-\frac{k}{n}))^{1/2})$

1er tratamiento $\Rightarrow$ Si el tratamiento está bien establecido entonces quiere decir que conocemos la distribución de población exacta y su media $\Rightarrow$

La media de la distribución es $\Rightarrow k = 80$ (número de pacientes que se curan)

2º tratamiento: Sabemos que la distribución de población es gaussiana pero no conocemos su media $\Rightarrow$ tenemos que estimar la media al 95% de confianza

- Distribución de población $P(x|k) = \dfrac{1}{(2\pi k(1-\frac{k}{n}))^{1/2}} \exp\left[-\dfrac{(x-k)}{2k(1-\frac{k}{n})^{1/2}}\right]$
- Estimador $\hat{k} = x$ (nº de pacientes que se curan al hacer cada test)
- Distribución muestral: $P(\hat{k}|k) = \dfrac{1}{(2\pi k(1-\frac{k}{n}))^{1/2}} \exp\left[-\dfrac{(\hat{k}-k)}{2k(1-\frac{k}{n})}\right]$

Sabiendo que $\hat{k}_{obs} = 85$ tenemos que hallar el intervalo de confianza para k que es $[k_-, k_+]$. Para que los cálculos sean más sencillos hacemos una aproximación en $P(\hat{k}|k) \Rightarrow (k(1-\frac{k}{n}))^{1/2} \sim [85(1-0.85)]^{1/2} = [85 \cdot 0.15]^{1/2} \simeq 3.6$

Si definimos $\sigma_{\hat{k}} = 3.6$, entonces la fórmula que hay que utilizar es:

$$\left.\begin{array}{l} k_+ = \hat{k}_{obs} + \sigma_{\hat{k}}\, \Phi^{-1}[1-\alpha] \\ k_- = \hat{k}_{obs} - \sigma_{\hat{k}}\, \Phi^{-1}[1-\beta] \end{array}\right\}$$

Nos piden el intervalo central de confianza $\Rightarrow \alpha = \beta$

Nivel de confianza del 95% $\Rightarrow 1-2\alpha = 0.95 \Rightarrow 0.05 = 2\alpha \qquad \alpha = 0.025 \Rightarrow 1-\alpha = 0.975$

$$k_\pm = 85 \pm 3.6\, \underbrace{\Phi^{-1}[0.975]}_{\text{usar tabla} \rightarrow 1.96} = 85 \pm 3.6 \times 1.96 \Rightarrow \begin{array}{l} k_+ = 92 \\ k_- = 78 \end{array}$$

Intervalo de confianza $[78, 92] \Rightarrow$ Podemos afirmar con un nivel de confianza del 95% que en aplicaciones futuras del 2º tratamiento el número mínimo de pacientes curados será 78 y el número máximo de pacientes curados será 92. Así que no podemos excluir con un nivel de confianza del 95% que el segundo tratamiento sea más efectivo que el primero porque el límite inferior del intervalo de confianza es menor que el número medio de pacientes que se curan con el primer tratamiento.

$\boxed{7}$ Queremos comparar la efectividad de dos tratamientos. En concreto, N_A enfermos han seguido el tratamiento A. De entre estos, S_A se han curado, mientras que no así el resto $(N_A - S_A)$. Correspondientemente N_B enfermos han seguido el tratamiento B, curándose S_B de entre ellos. La efectividad del tratamiento i-ésimo la expresamos por una probabilidad p_i: la probabilidad de que el enfermo se cure, dado que ha seguido el tratamiento i. Por tanto, queremos estimar p_B y p_A.

a) ¿Son los estadísticos S_B/N_B y S_A/N_A estimadores insesgados de p_B y p_A respectivamente ?

b) Escriba el intervalo de confianza de 95 % para cada p_i.

c) Calcule el error estándar de los estimadores del apartado a).

Sugerencia: consiga la distribución de probabilidad correspondiente a que S_A (S_B) sanen de entre N_A (N_B) que han seguido el tratamiento, considere en este caso también que $N_A, N_B >> 1$.

Definimos $\quad p_A \equiv \dfrac{S_A}{N_A} \qquad p_B \equiv \dfrac{S_B}{N_B}$

$\Rightarrow$ Distribución de probabilidad de que S_A sanen : Binomial $B(N_A, p_A)$

$$p(S_A) = \binom{N_A}{S_A} p_A^{S_A} (1-p_A)^{N_A - S_A} \qquad S_A = 0, 1, 2 \ldots N_A$$

$\Rightarrow$ Distribución de probabilidad de que S_B sanen : Binomial $B(N_B, p_B)$

$$p(S_B) = \binom{N_A}{S_B} p_B^{S_B} (1-p_B)^{N_B - S_B} \qquad S_B = 0, 1, 2 \ldots N_B$$

Si $N_A, N_B >> 1$ entonces aproximamos las binomiales por normales

$$B(N_A, p_A) \rightarrow N\left[N_A p_A, \left(N_A p_A (1-p_A)\right)^{\frac{1}{2}}\right]$$

$$B(N_B, p_B) \rightarrow N\left[N_B p_B, \left(N_B p_B (1-p_B)\right)^{\frac{1}{2}}\right]$$

Por tanto tenemos que

$$S_A \sim N\left[\underbrace{N_A p_A}_{\mu_A}, \underbrace{\left(N_A p_A (1-p_A)\right)^{\frac{1}{2}}}_{\sigma_A}\right], \quad S_B \sim N\left[\underbrace{N_B p_B}_{\mu_B}, \underbrace{\left(N_B p_B (1-p_B)\right)^{\frac{1}{2}}}_{\sigma_B}\right]$$

Distribuciones de población para cada tratamiento:

$$P(S_A \mid \mu_A, \sigma_A) = \frac{1}{\sigma_A\sqrt{2\pi}} \exp\left[-\frac{(S_A-\mu_A)^2}{2\sigma_A^2}\right]$$

$$P(S_B \mid \mu_B, \sigma_B) = \frac{1}{\sigma_B\sqrt{2\pi}} \exp\left[-\frac{(S_B-\mu_B)^2}{2\sigma_B^2}\right]$$

Variables aleatorias:

S_A, S_B = número de pacientes que se curan

a) Supongamos que ahora tomamos los estadísticos $\frac{S_B}{N_B}$, $\frac{S_A}{N_A}$ como estimadores de p_B y p_A respectivamente. Definimos los estimadores

$$\hat{p}_B = \frac{S_B}{N_B}, \quad \hat{p}_A = \frac{S_A}{N_A}$$

Calculamos la distribución muestral de $\hat{p}_B$: hacer el cambio de variables aleatorias $S_B = \hat{p}_B N_B$

$dS_B = N_B \, d\hat{p}_B \Rightarrow$

$$P(\hat{p}_B \mid \mu_B, \sigma_B) = \frac{N_B}{\sigma_B\sqrt{2\pi}} \exp\left[-\frac{(N_B\hat{p}_B - \mu_B)^2}{2\sigma_B^2}\right]$$

$$= \frac{1}{\frac{\sigma_B}{N_B}\sqrt{2\pi}} \exp\left[-\frac{\left(\hat{p}_B - \frac{\mu_B}{N_B}\right)^2}{2\frac{\sigma_B^2}{N_B^2}}\right] = \frac{1}{p_B(1-p_B)\sqrt{2\pi}} \exp\left[-\frac{(\hat{p}_B - p_B)^2}{2p_B^2(1-p_B)^2}\right]$$

$$\Rightarrow \hat{p}_B \sim N[p_B, p_B(1-p_B)]$$

donde utilizamos que

$$\mu_B = N_B p_B$$
$$\sigma_B = (N_B p_B (1-p_B))^{\frac{1}{2}}$$

Análogamente se prueba que $\hat{p}_A \sim N[p_A, p_A(1-p_A)]$. Usamos las distribuciones muestrales para calcular $E(\hat{p}_B)$, $E(\hat{p}_A)$. El cálculo es directo porque son distribuciones normales

$$\left.\begin{aligned} E(\hat{p}_B) &= p_B \\ E(\hat{p}_A) &= p_A \end{aligned}\right\} \Rightarrow \text{no hay sesgo.}$$

b) Calculamos el intervalo de confianza $[p_A^-, p_A^+]$ para p_A

(el cálculo para p_B es análogo).

$$\int_{(\hat{p}_A)_{obs}}^{\infty} \frac{1}{p_A^-(1-p_A^-)\sqrt{2\pi}} \exp\left[-\frac{(\hat{p}_A - p_A^-)^2}{2(p_A^-)^2(1-p_A^-)^2}\right] d\hat{p}_A = \beta$$

$$\int_{-\infty}^{(\hat{p}_A)_{obs}} \frac{1}{p_A^+(1-p_A^+)\sqrt{2\pi}} \exp\left[-\frac{(\hat{p}_A - p_A^+)^2}{2(p_A^+)^2(1-p_A^+)^2}\right] d\hat{p}_A = \alpha$$

El cálculo es el mismo que hacemos para una distribución normal pero sustituyendo $\sigma_{\hat{q}}$ por $p_A^{\pm}(1-p_A^{\pm})$

$$\left.\begin{aligned} \Phi\left[\frac{(\hat{p}_A)_{obs} - p_A^+}{p_A^+(1-p_A^+)}\right] &= \alpha \\ \Phi\left[\frac{(\hat{p}_A)_{obs} - p_A^-}{p_A^-(1-p_A^-)}\right] &= 1-\beta \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} (\hat{p}_A)_{obs} - p_A^+ &= p_A^+(1-p_A^+)\,\Phi^{-1}[\alpha] \\ (\hat{p}_A)_{obs} - p_A^- &= p_A^-(1-p_A^-)\,\Phi^{-1}[1-\beta] \end{aligned}\right.$$

Para simplificar el cálculo hacemos la siguiente aproximación en el sistema:

$$p_A^+(1-p_A^+) \sim (\hat{p}_A)_{obs}\left(1-(\hat{p}_A)_{obs}\right)$$

Es la misma aproximación que se hizo en el problema 6

Resolvemos ahora el sistema suponiendo que $\alpha = \beta \Rightarrow$ usar $\Phi^{-1}[1-\alpha] = -\Phi^{-1}[\alpha]$

$$\boxed{\begin{aligned} p_A^+ &= (\hat{p}_A)_{obs} - (\hat{p}_A)_{obs}\left(1-(\hat{p}_A)_{obs}\right)\Phi^{-1}[\alpha] \\ p_A^- &= (\hat{p}_A)_{obs} + (\hat{p}_A)_{obs}\left(1-(\hat{p}_A)_{obs}\right)\Phi^{-1}[\alpha] \end{aligned}}$$

Para un intervalo central de confianza del 95% $1-2\alpha = 0.95 \quad 0.05 = 2\alpha$

$\alpha = 0.025 \Rightarrow$ usando la tabla $\Phi^{-1}[\alpha] = -1.960$.

c) el error estandard de los estimadores se saca a partir de las distribuciones muestrales, que son normales:

$$\hat{p}_A \sim N[p_A, p_A(1-p_A)] \Rightarrow \sigma_{\hat{p}_A} = p_A(1-p_A)$$

$$\hat{p}_B \sim N[p_B, p_B(1-p_B)] \Rightarrow \sigma_{\hat{p}_B} = p_B(1-p_B)$$