Auxiliar 1
CC50Q

Prof: Pedro Ortega C. `<peortega@dcc.uchile.cl>`
Aux: Francisco Claude F. `<fclaude@dcc.uchile.cl>`

2 de agosto de 2005

Problema 1

Una caja contiene $2n$ helados, n de los cuales son de naranja y n de frutilla. De un grupo de $2n$ personas m prefieren el helado de naranja ($0 < m < n$), s prefieren el de frutilla ($0 < s < n$) y el resto no tiene preferencia. Los helados se distribuyen al azar, con los participantes ordenados de forma que primero están los que prefieren naranja, luego los que prefieren frutilla y finalmente el resto. ¿ Cuál es la probabilidad de que se respeten las preferencias de todas las personas?

Solución Problema 1

Se tienen $2n$ helados, la probabilidad de que los que prefieren helado de naranja reciban lo que esperan es:

$$\frac{n}{2n}\frac{n-1}{2n-1}\frac{n-2}{2n-2}\cdots\frac{n-m+1}{2n-m+1}$$

Luego la probabilidad de que los que prefieren helado de frutilla reciban los que esperan es:

$$\frac{n}{2n-m}\frac{n-1}{2n-m-1}\frac{n-2}{2n-m-2}\cdots\frac{n-m+1}{2n-m-s+1}$$

Finalmente mutiplicando los dos valores y agrupando tenemos:

$$\frac{n!n!}{(2n)!}\frac{(2n-m-s)!}{(n-m)!(n-s)!}$$

Por último se pueden reagrupar los términos:

$$\frac{\frac{(2n-m-s)!}{(n-m)!(n-s)!}}{\frac{(2n)!}{n!n!}} = \frac{\binom{2n-m-s}{n-m}}{\binom{2n}{n}}$$

Problema 2

Hay tres puertas, A, B y C. Un premio ha sido escondido tras una de las puertas. Ud. puede escoger una puerta. Inicialmente la puerta que Ud. escogió no es abierta. A cambio, el anfitrión del juego abrirá una de las dos restantes, *de manera de no revelar la puerta con el premio*. Tras ésto, Ud. puede cambiar su elección o mantenerla igual.

Si su elección inicial es la puerta A y el anfitrión abre la puerta C, mostrando que ahí no esta el premio. ¿ Conviene cambiar la elección ?

Solución Problema 2

Inicialmente sabemos que el premio (P) se esconde en A,B o C. Como inicialmente no tenemos información alguna sobre tras que puerta se esconde el premio, podemos asumir una distribución uniforme:

$$\begin{aligned} P(P=A) &= \frac{1}{3} \\ P(P=B) &= \frac{1}{3} \\ P(P=C) &= \frac{1}{3} \end{aligned}$$

Luego de nosotros elegir la puerta A, sabemos que la probabilidad de que esta esconda el premio es $\frac{1}{3}$. Pero luego de que esto ocurra, el anfitrión abre una puerta, sabiendo que esta no esconde el premio, por lo que tenemos una ganancia de información, la cúal se puede plantear de la siguiente forma:

$$\begin{aligned} P(Anf=B|P=A) &= 1/2 \\ P(Anf=C|P=A) &= 1/2 \\ P(Anf=B|P=B) &= 0 \\ P(Anf=C|P=B) &= 1 \\ P(Anf=B|P=C) &= 1 \\ P(Anf=C|P=C) &= 0 \end{aligned}$$

Donde Anf corresponde al la puerta abierta por el anfitrión. Como se puede ver, el hecho de saber que puerta abre el anfitrión esta relacionada con la puerta que nosotros elegimos y la que en realidad esconde el premio. Esto nos permitirá determinar cúal de las dos puertas que nos queden será la más probable. Usando Bayes en cada caso, obtenemos:

$$
\begin{aligned}
P(P = A|Anf = C) &= \frac{(1/2)(1/3)}{P(Anf = C)} \\
P(P = B|Anf = C) &= \frac{(1)(1/3)}{P(Anf = C)} \\
P(P = C|Anf = C) &= 0
\end{aligned}
$$

Tenemos que $P(Anf = C)$ es una constante que debe normalizar nuestra distribución, por lo que podemos usar esta propiedad para obtenerla:

$$
\begin{aligned}
1 &= \frac{1}{P(Anf = C)} \left(\frac{1}{6} + \frac{2}{6} \right) \\
P(Anf = C) &= \frac{3}{6} = \frac{1}{2}
\end{aligned}
$$

Finalmente calculamos las probabilidades a posteriori:

$$
\begin{aligned}
P(P = A|Anf = C) &= \frac{1}{3} \\
P(P = B|Anf = C) &= \frac{2}{3} \\
P(P = C|Anf = C) &= 0
\end{aligned}
$$

Por último podemos concluir que lo más conveniente es cambiarse de puerta luego de ver que el anfitrión abrió la puerta C y esta no escondía un premio.

Problema 3

Se tiene dos cajas (A y B) de tal forma que la caja A contiene 8 fichas blancas y 2 negras, en tanto B contiene X fichas blancas y $10 - X$ negras. Un juego consiste en lanzar un dado (a costo de 2 UM), si sale mayor a 4 se saca una ficha de la caja A, en caso contrario se saca de la caja B. Si la ficha obtenida es blanca se obtiene un ingreso de 4 UM, si es negra 0 UM.

1. ¿ Cuál es la probabilidad de obtener un ficha blanca ?
2. ¿ Cuanto debe valer X para que el juego le convenga en promedio a su dueño ?

Solución Problema 3

Para poder determinar X, debemos primero determinar la probabilidad de que el jugador obtenga una ficha blanca. Esta se calcula sabiendo que la probabilidad de obtener un número mayor a 4 con un dado es $\frac{1}{3}$, luego sabemos que:

$$\begin{aligned} P(F = Blanca) &= \frac{1}{3} \times \frac{8}{10} + \left(1 - \frac{1}{3}\right) \times \frac{X}{10} \\ P(F = Blanca) &= \frac{8}{30} + \frac{2X}{30} \end{aligned}$$

Para hacer que el juego sea conveniente al dueño, planteamos que en promedio el jugador pierda, es decir:

$$E[4 \times P(F = Blanca) - 2] < 0$$

$$\begin{aligned} 4 \times \frac{2X + 8}{30} - 2 &< 0 \\ 2X &< \frac{2 \times 30}{4} - 8 \\ 2X &< 15 - 8 \\ X &< \frac{7}{2} = 3{,}5 \end{aligned}$$

Finalmente concluimos que $X = 3$ basta para que en promedio al dueño del juego le sea conveniente.

Problema 4

Se tiene un dispositivo que al emplearlo puede producir la muerte o la no-muerte con una cierta probabilidad. Realizando simulaciones, se observa que, de 300 experimentos, en 250 de ellos se produjo la muerte, y en los casos restantes, no. Se piensa que la probabilidad de ocurrir muerte es inferior o igual a 0.8 con un nivel de confianza de 0.05.

Argumente la válidez de esta hipotesis desde una perspectiva bayesiana.

Solución problema 4

Procedimiento

Para resolver este problema, debemos calcular $f(p|x_1, x_2, ..., x_{300})$, donde inicialmente p es desconocido, utilizando el método de Bayes, el procedimiento para obtener esta función es el siguiente:

- Se asume una cierta distribución de probabilidad para el parámetro p
- Se calcula $f(x_1, x_2, ..., x_{300}, p) = f(x_1, x_2, ..., x_{300}|p)f(p)$
- Utilizando el item anterior obtenemos la distribución marginal $g(x_1, x_2, ..., x_{300})$
- Finalmente calculamos $f(p|x_1, x_2, ..., x_{300}) = \frac{f(x_1,x_2,...,x_{300},p)}{g(x_1,x_2,...,x_{300})}$

Teniendo esta información, podemos determinar la probabilidad de que p tome un valor menor o igual a 0.8, para esto basta calcular:

$$\int_0^{0,8} f(p|x_1, x_2, ..., x_{300})dp$$

El intervalo de confianza no es necesario para la conclusión, de todas formas si se quisiera calcular, se debiese estimar p y con este valor estimado, se podría calcular un intervalo de Bayes para el parámetro, el cuál acepta un cierto error.

Supuestos y datos previos

Para no asumir una cierta tendencia del valor de p, asumiremos que tiene una distribución $f(p)$ uniforme determinada por $f(p) = 1, 0 \leq p \leq 1$
Además podemos modelar la distribución de $f(x_1, x_2, ..., x_{300}|p)$ representandola mediante la probabilidad de n muertes, ie, $P(n) = \binom{300}{n} p^n (1-p)^{300-n}$

Cálculo de $f(x_1, x_2, ..., x_{300}, p)$

Sabemos que $f(x_1, x_2, ..., x_{300}, p) = f(x_1, x_2, ..., x_{300}|p)f(p)$, mediante esta información y la descrita en el punto anterior, despejamos

$$f(x_1, x_2, ..., x_{300}, p) = 1\binom{300}{n} p^n (1-p)^{300-n}$$

[1]

Cálculo de $g(x_1, x_2, ..., x_{300})$

$$g(x_1, x_2, ..., x_{300}) = \int_0^1 \tbinom{300}{n} p^n (1-p)^{300-n} dp$$

Para este caso en particular tenemos:

$$g(x_1, x_2, ..., x_{300}) = \int_0^1 \tbinom{300}{250} p^{250} (1-p)^{50} dp$$

La resolución de esta integral no es trivial, por el momento es suficiente como información, ya que la parte númerica podremos resolverla al final.

Cálculo de $f(p|x_1, x_2, ..., x_{300})$

$$f(p|x_1, x_2, ..., x_{300}) = \frac{f(x_1, x_2, ..., x_{300}, p)}{g(x_1, x_2, ..., x_{300})}$$

Reemplazando los datos obtenidos anteriormente, obtenemos:

$$f(p|x_1, x_2, ..., x_{300}) = \frac{f(x_1, x_2, ..., x_{300}, p)}{\int_0^1 \tbinom{300}{250} p^{250} (1-p)^{50} dp}$$

Calculando la probabilidad de $p \leq 0{,}8$

Para el cálculo de esta probabilidad, basta integrar la función obtenida anteriormente entre 0 y 0.8, por lo que nuestro problema se resume ahora al cálculo de:

$$\frac{\int_0^{0,8} f(x_1, x_2, ..., x_{300}, p) dp}{\int_0^1 \tbinom{300}{250} p^{250} (1-p)^{50} dp}$$

Reemplazando tenemos:

$$\frac{\int_0^{0,8} f(x_1, x_2, ..., x_{300}, p) dp}{\int_0^1 \tbinom{300}{250} p^{250} (1-p)^{50} dp} = \frac{\int_0^{0,8} \tbinom{300}{250} p^{250} (1-p)^{50} dp}{\int_0^1 \tbinom{300}{250} p^{250} (1-p)^{50} dp}$$

Con la ayuda de maple, obtenemos el siguiente resultado:

$$\frac{\int_0^{0,8} \tbinom{300}{250} p^{250} (1-p)^{50} dp}{\int_0^1 \tbinom{300}{250} p^{250} (1-p)^{50} dp} = 0{,}07881194754$$

[1] Aqui se ve un abuso de notación, que utilizaré por motivos de simplicidad, el cuál asume que la probabilidad de $x_1, x_2, ..., x_{300}$ es la probabilidad de las 250 muertes contando todas sus permutaciones, no respetando $x_1, x_2, ..., x_{300}$ como un conjunto ordenado.

Conclusiones

Del cálculo anterior, tenemos que $P(p \leq 0{,}8) = 0{,}0788$ de esto deducimos claramente que es mucho mas probable encontrar a p sobre 0.8 y por ende la hipotesis tiene una credibilidad muy baja. Con una probabilidad asi, claramente se rechaza la hipotesis y se asume $p \geq 0{,}8$

Cabe destacar que no se asumió ningun valor para p, se intento concluir tratando de asumir lo mínimo, en este caso eso fue pensar que cualquier valor de p entre 0 y 1 era igual de válido. Si se planteará que p tiene alguna otra distribución, quiza pensando algo que tuviese una media en $\frac{250}{300}$ se podrían obtener resultados diferentes. Con la información contenida en el enunciado, no es posible hacer algún supuesto de este tipo, por lo que se debe partir de la base que p no esta restringido ni tiene tendencia a tomar ciertos valores (salvo que debe estar entre 0 y 1, por definición de probabilidad).

En la solución propuesta, seguiendo métodos clásicos, la hipotesis no se podía rechazar, lo que no significa que esta sea correcta, simplemente que la región de rechazo era un poco más amplia de lo que se necesitaba para rechazar. Al probar utilizando métodos Bayesianos, obtuvimos que la hipotesis si se rechaza o mas bien, es muy poco probable su veracidad. Dado además el resultado obtenido de la forma clásica, donde el valor $\frac{250}{300}$ se encuentra muy cercano a la región de rechazo es posible creer que el cálculo obtenido mediante la resolución Bayesiana es igualmente válido y posiblemente más preciso para este caso.