4. Modelos de distribución teórica en estadística

Modelización de Sistemas Ambientales (07M1)
Dpto. de Ecología e Hidrología y Dpto. Geografía
Facultad de Biología
Universidad de Murcia

Curso 2006–07

Índice

1. Introducción

El modo de observar un proceso determina las variables que han de medirse, las hipótesis de trabajo definen que comportamiento esperamos para tales variables.

Estudiaremos un caso de distribución aleatoria de individuos sobre un gradiente, es decir, donde no hay efecto de tal gradiente en la distribución de los organismos. Por lo tanto cualquier región del gradiente tiene la misma probabilidad de ser ocupada por un individuo. También puede expresarse como: "dos regiones iguales en extensión tendrán la misma densidad teórica".

Podemos simular la posición de los individuos considerando como variables las coordenadas que indican su ubicación en el espacio y simulando estas coordenadas según una distribución aleatoria.

Analizaremos el comportamiento de las variables "observables" en comparación con las variables que producen la simulación.

2. Distribución espacial de organismos (3 horas)

EJERCICIOS

1. Considerando que los individuos se situar sobre un gradiente, de longitud 100 y que lo hacen de forma independiente puede generarse la ubicación de 150 individuos mediante la función `runif()`:

   ```
   x<-runif(150,0,100)
   plot(x,rep(1,150),cex=0.3,xlab="Gradiente",ylab="")
   abline(v=0:100,col=3)
   ```

   Los resultados se representan mediante la función `plot()`, el eje de abscisas indica la posición en el gradiente, el eje de ordenadas no tiene más interés que el de facilitar la representación. La opción `cex=0.3` permite representar los puntos de menor tamaño (cuando no se indica lo contrario `cex` toma valor el valor por defecto, este es 1). La líneas verdes delimitan cien regiones en el área de trabajo a las que llamaremos unidades de muestreo; es este caso el muestreo es sistemático e intensivo.

   Considerando la variable "*individuos por unidad de muestreo*"? ¿Cuántas unidades de muestreo no presentan individuos? ¿Cuántas presentan sólo un individuo? ¿Cuál es el máximo número de individuos por unidad de muestreo? ¿Cuál es la densidad media? ¿Puede calcularse rápidamente?

   Si los individuos hubiesen presentado patrones de selección, con la preferencia de una de zona dada ¿Cómo podría cambiar la densidad media esperada?

2. Desde un punto de vista teórico la densidad en cada uno de los intervalos de igual longitud en que se divide el gradiente debería proporcionar un número de individuos aproximadamente constante, y de un 100–ávo del número de individuos simulado. Mediante:

```
table(trunc(x)+1)
```

conseguimos, de forma rápida, contar cuantos individuos caen en cada una de las 100 partes en las que dividimos el gradiente.

¿Qué hace la función `trunc()`? ¿Cuál es el valor más pequeño que devuelve `trunc(x)`? ¿Qué significan los valores que devuelve la función `table()`? ¿Cuántos individuos hay en la unidad muestreo número 1? ¿y en la 2? ¿Cuantas unidades de muestreo no presentan ni un sólo individuo? ¿Cuantas muestras se han tomado?

3. Si utilizamos la tabla como la variable que describe al número de individuos por unidad de muestreo, podemos obtener algunos resultados interesantes:

   ```
   table(trunc(c(x,0:99))+1)-1->nipum
   table(nipum)
   mean(nipum)
   ```

   Se ha utilizado un artificio para simplificar la inclusión de unidades muestrales vacías. Se trata de asegurar un individuo en cada unidad de muestreo y posteriormente en la tabla reducir una unidad el número de observados.

   ¿Qué significa la nueva tabla? ¿Qué indica la media? ¿Qué valor para la media de `ipum` cabía esperar, conocidos el número total de individuos simulados y el tamaño muestral? La nueva variable es ¿continua o discreta? su comportamiento ¿es normal o sigue otra distribución de probabilidad? ¿Cuál es el tamaño muestral? ¿Cuál es la frecuencia con la que aparecen 3 individos por unidades de muestreo? ¿Y con 3 individuos o más?

4. Supongamos que la variable `nipum` sigue una distribución de Poisson. Utilizando la función `dpois()` para calcular la frecuencia teórica para los distintos valores de la variable 0, 1, 2, . . . :

   ```
   round(dpois(0:10,mean(nipum))*100)
   ```

   ¿Cómo pueden compararse los resultados obtenidos del muestreo con lo teóricos?

5. Una de las predicciones de la distribución de Poisson es que:

   $$P(X = 0) = e^{-\lambda}$$

   siendo λ la media poblacional de la variable X y por tanto cuando se conoce la frecuencia de unidades de muestreo vacías ($X = 0$) puede estimarse la densidad media utilizando:

   $$\lambda = -log(P(X = 0))$$

   ¿Ocurre esto en nuestro caso? ¿Es una prueba a favor o en contra para ratificar la hipótesis de que la variable `nipum` sigue distribución de Poisson? Si repetimos la simulación: ¿Cambia la frecuencia de unidades de muestreo vacías? ¿Estimamos una densidad media muy distinta a la real?

6. Considerando los resultados obtenidos de las distintas simulaciones podemos afirmar que: *la distribución de la variable número de individuos por unidad de muestreo sigue una distribución de Poisson con parámetro λ igual a la media de la variable, que puede aproximarse a la media muestral*.

   ¿Cómo se expresarían estos resultados para apoyar la hipótesis?

7. Un argumento en contra puede basarse en la siguiente afirmación: *la densidad media puede afectar al resultado*, es decir, para densidades distintas a la simulada esta situación varia considerablemente de la predicha por la distribución de Poisson. El argumento en que se basa esta afirmación es el siguiente: "*Con densidades muy altas no se encontrarán unidades de muestreo vacías y por tanto no se cumplirá: $\lambda = -log(P(X = 0))$*".

   Para definir poder realizar cómodamente la simulación con distinto número de individuos utilizaremos la función `experimento()`:

```
experimento<-function(nis=150){
   x<-runif(nis,0,100)
   plot(x,rep(1,nis),cex=0.3)
   abline(v=0:100,col=3)
   table(trunc(c(x,0:99))+1)-1->nipum
   table(nipum)->observadas
   mean(nipum)
   as.numeric(names(table(nipum)))->vov
   round(dpois(vov,mean(nipum)),2)->esperadas
   observadas
   esperadas*100
   chisq.test(as.vector(observadas),esperadas)
   }
```

utilizando esta es fácil repetir el experimento considerando un número variable de individuos totales en el gradiente y por tanto una densidad media variable. Las variables utilizadas son: `nis`, o número de indididuos simulados; `nipum`, número de individuos por unidad de muestreo; `observadas`, frecuencias observadas para la variable número de individuos por unidad de muestreo; `esperadas`, probabilidad de cada uno de los valores de la variable.

¿Se acepta el test de χ^2 de bondad de ajuste? ¿Qué implica este resultado en relación con la hipótesis de trabajo?

¿Cómo se diseñaría un experimento para aceptar o rechazar el efecto de la densidad media?

¿Qué resultados proporciona el experimento? ¿Avalan o rebaten la hipótesis?

3. Para entregar

Considerando los resultados obtenidos para discutir sobre: "*La adecuación del uso de la distribución de Poisson para predecir la distribución de frecuencias de la variable número de individuos por unidad de muestreo cuando la distribución espacial de los individuos sigue una distribución uniforme*". Escribir muy breve y concisamente: la metodología, los resultados, la discusión y las conclusiones a las que se llega dado el experimento propuesto.

Para una discusión sobre el significado de la distribución de Poisson en la variable número de individuos por unidad de muestreo puede consultarse Bolanos[1] o en Santiago Clavijo[2].

Para más detalles sobre la distribución de Poisson en la página de Rodríguez Riotorto[3]:

[1] http://biologia.ucr.ac.cr/profesores/Federico%20Bolanos/Ecologia%20General/Distribucion%20Espacial.pdf
[2] http://www.plagas-agricolas.info.ve/doc/html/manejo_plagas/capitulo_4.html
[3] http://www.telefonica.net/web2/biomates/secundaria/distprob/index.htm