Experimentos de Monte Carlo

Walter Sosa-Escudero

Test de Breusch-Pagan de heterocedasticidad:

$$LM = \frac{1}{2}\, SCE_{g,z} \sim \chi^2(p-1)$$

$g =$ residuos al cuadrado, z, variables explicativas de la heterocedasticidad. Esta es una aseveracion asintotica. Que sucede en muestras finitas?

- Resultados exactos: pueden ser analiticamente complejos, sino imposibles.
- Experimento de Montecarlo: procedimiento en donde una magnitud de interes es aproximada por un procedimiento aleatorio.

Objetivo: estudiar las propiedades de muestra finita de un procedimiento estadistico a traves de un experimento computacional.

Un ejemplo mas concreto:

$$Y_i \sim (\mu, \sigma^2), i = 1, \dots, n$$

Consideremos:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2,$$

consistente para σ^2. Es bien conocido que:

$$E(\hat{\sigma}^2) = \sigma^2 \frac{n-1}{n}$$

de modo que el estimador es *sesgado*. A este resultado se puede llegar analiticamente, pero podriamos haberlo aproximado con el siguiente experimento computacional.

1. Fijamos un tamaño de muestre, $n = 20$
2. Generamos J muestras de tamaño 20, de la distribucion $N(0,1)$. En cada paso computamos $\hat{\sigma}_j^2$, $j = 1, \ldots, J$.
3. Computamos $\frac{1}{K}\sum_{k=1}^{K} \hat{\sigma}_k^2$ y vemos si es cercano a 1.

Preguntas:

- Cuan informativo es este procedimiento del resultado que buscamos.
- En que medida el resultado depende de parametros especificos (n, σ^2, μ) fijados en el experimento?
- Hasta que punto los resultados son extrapolables a otros contextos?

Supongamos que nos interesa computar:

$$\theta = \int_0^1 g(x)dx$$

en donde $g(x)$ es cualquier funcion real. Este problema es completamente deterministico.

Si $U \sim$ U(0,1), $f(u) = 1[0 \leq u \leq 1]$, entonces:

$$\theta = E\ g(U)$$

Si $U_j \sim U(0,1)$ iid , $j = 1, \ldots, J$ entonces, $g(U_j), j = 1, \ldots, J$ es iid con $Eg(U_j) = \theta$.

De acuerdo a la ley de grandes numeros:

$$\frac{1}{J}\sum_{j=1}^{J} g(U_j) \xrightarrow{p} E(g(U)) = \theta$$

Entonces, podemos *aproximar* θ generando una muestra grande de $U_j \sim U(0,1)$ y tomar como aproximacion:

$$\tilde{\theta} = \frac{1}{J}\sum_{j=1}^{J} g(U_j)$$

Para J suficientemente grande, este procedimiento no deberia diferir del procedimiento de integracion analitica.

A este mecanismo, de reemplazar una integral por una esperanza y aproximarla por un mecanismos aleatorio, se lo llama *integracion por Monte Carlo*.

- $\theta = \int_a^b g(x)$ requiere un simple cambio de variables.
- Caso multidimensional: $g(x_1, \ldots, x_n)$ y queremos computar:

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 g(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

Es facil observar que:

$$\theta = E[g(U_1, \ldots, U_n)]$$

en donde $U_1, \ldots, U_n$ son $U(0,1)$ iid.

- Generar J conjuntos de n numeros iid $U(0,1)$

$$\begin{array}{ccc} U_{11} & \cdots & U_{n1} \\ U_{12} & \cdots & U_{n1} \\ \vdots & & \vdots \\ U_{1J} & \cdots & U_{nJ} \end{array}$$

y computar:

$$\tilde{\theta} = \frac{1}{J} \sum_{j=1}^{J} g(U_{1j}, \ldots, U_{nj})$$

- El uso de la distribucion uniforme es arbitrario. Notar que:

$$\theta = \int g(x)dx = \int \frac{g(x)}{f(x)} f(x)dx = E(z(X))$$

en donde $z(X) \equiv g(X)/f(X)$ y $X \sim f(x)$. Entonces, el procedimiento consiste en tomar

$$X_j, j = 1, \ldots, n, iid \sim f(x)$$

y computar:

$$\tilde{\theta} = \frac{1}{J} \sum_{j=1}^{J} z(X_i)$$

Notar que $f(x)$, entonces, puede ser elegida arbitrariamente.

Ejemplo: $\hat{\sigma}^2$ y S^2

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

queremos computar $E(\hat{\sigma}^2)$ y $E(S^2)$. Lo haremos para un n dado. En nuestra notacion anterior:

$$\begin{aligned} E(\hat{\sigma}^2) &= E(g(x_1, \ldots, x_n)) \\ &= \int \ldots \int g(x_1, \ldots, x_n) f(x_1) \cdots f(x_n) dx_1 \cdots f(x_n) \end{aligned}$$

en donde $f(x_i)$ es la verdadera distribucion de los datos. Es importante notar que $E(\hat{\sigma}^2)$ *no* es una variable aleatoria.

Recordar que analiticamente se puede probar que si X_i, iid con $V(X_i) = \sigma^2$, entonces:

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

y que este resultado es cierto para cualquier media poblacional y para cualquier distribucion.

El procedimiento a seguir sera el siguiente:

- Fijamos $n = N$
- Suponemos $X_i \sim (\mu_0, \sigma_0^2)$ μ_0 puede ser elegido arbitrariamente, y la distribucion de X_i tambien. Fijamos un valor para σ_0^2.
- Generamos J muestras iid de tamaño N: $X_{1j}, X_{2j}, \ldots, X_{Nj}, j = 1, \ldots, J$.
- Para cada muestra computamos:

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2$$

- Aproximamos $E(\hat{\sigma}^2)$ usando:

$$\frac{1}{J} \sum_{j=1}^{J} \hat{\sigma}_j^2$$

Numeros aleatorios

Las computadoras no pueden generar 'numeros aleatorios', sino numeros 'pseudo-aleatorios': sucesion deterministica de numeros que tiene las mismas propiedades estadisticas que los numeros aleatorios.

Ejemplo: generador congruencial: genera numeros $U(0,1)$

$$x_n = (ax_{n-1} + c) \text{ mod } m$$

en donde a y c son dos constantes y m es usualmente una constante muy grande.

Metodo de transformacion inversa: permite generar numeros aleatorios a partir de un numeros aleatorios uniformes.

Resultado: si $U \sim Unif(0,1)$ y $F(z)$ es la funcion de distribucion acumulada de una variable continua:

$$X = F^{-1}(U) \sim F$$

Prueba: $U \sim Unif(0,1) \Rightarrow Pr(U < x) = x$

$$\begin{aligned} P(X \leq x) &= P(F^{-1}(U) \leq x) \\ &= P(F(F^{-1}(U)) \leq F(x)) \\ &= P(U \leq F(x)) \\ &= F(x) \end{aligned}$$

ya que U es uniforme.

Esto implica el siguiente proceso para generar datos aleatorios de una variable aleatoria con distribuion F:

1. Generar U_i de la distribucion uniforme
2. Tomer $X_i = F^{-1}(U_i)$

Diseño de experimentos de montecarlo

Un experimento de montecarlo requiere:

1. Un proceso generador de datos (DGP), definido para un conjunto de *parametros* en sentido amplio
2. Un estadistico de interes
3. Un objetivo del experimento, que usualmente consiste en alguna propiedad del estadistico de interes en relacion a los posibles parametros del modelo

Un ejemplo:

$$y_t = \alpha y_{t-1} + u_t, t = 1, \ldots, T \tag{1}$$

$$u_t \sim N(0, \sigma^2) \tag{2}$$

Un resultado estandar es que el estimador MCO de α es *sesgado*. Supongamos que nos interesa estudiar el sesgo a traves de un experimento de montecarlo.

En este contexto, (1) y (2) representan el DGP. Los parametros involucrados son α, σ^2 y T. El estadistico de interes es $\hat{\alpha} = \sum_{t=1}^{T} y_t y_{t-1} / \sum_{t=1}^{T} y_{t-1}^2$ y el objetivo del experimento es:

$$E(\hat{\alpha}|\alpha, \sigma^2, T)$$

- El objetivo del experimento es evaluar $E(\hat{\alpha})$ para distintos valores de α, σ^2 y T. Un 'parametro' adicional es la distribucion normal de los errores, tambien se podria variar esta distribucion.
- El *diseño experimental* se refiere a los valores de los parametros que se utilizaran para estudiar el problema.
- *Estadisticos pivotales:* un estadistico es *pivotal* con respecto a un parametro si su distribucion no depende de ese parametro. *Ejemplo:* $Y_i \sim N(\mu, \sigma^2)$, y consideremos:

$$\bar{Y} \sim N(\mu, \sigma^2/n), t \equiv \frac{\bar{Y} - \mu}{\sqrt{S^2}}$$

$\bar{Y}$ no es pivotal con respecto a σ^2 pero t si lo es.

- Un estadistico es *asintoticamente pivotal* con respecto a un parametro si asintoticamente su distribucion no depende de ese parametro.

  Ejemplo: $Y_i \sim \mu, \sigma^2)$ (no necesariamente normal). Lo unico que sabemos, por el TCL, es que t es asintoticamente pivotal La deteccion de parametros para los cuales el estadistico de interes es pivotal es crucial, porque?

El test de Breusch-Pagan

Setup:

$$y_i = x_i'\beta + u_i, \quad i = 1, 2, \ldots, N$$

- x_i vector de K variables explicativas, con constante.
- $u_i = N(0, \sigma_i^2)$
- $V(u_i) = \sigma_i^2 = h(z_i'\alpha)$
- z_i vector de p variables explicativas de la varianza (con constante)
- $h(.) =$ cualquier funcion positiva con 2 derivadas.
- Homocedasticidad: $H_0 : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0$
- e_i, residuos de OLS ignorando heterocedasticidad.
- $\hat{\sigma}^2 = (1/N) \sum e_i^2$. Estimador MV de σ^2 bajo H_0
- $g_t \equiv e_i^2 / \hat{\sigma}^2$

Test (Breusch-Pagan): Bajo H_0 :

$$LM_{BP} = \frac{1}{2}\, SCE_{g,z} \sim \chi^2(p-1)$$

$SCE_{g,z} = SCE$ de regresar g_i en z_i.
Mas especificamente, queremos evaluar la siguiente aseveracion. Consideremos el caso $p = 2$ (una sola variable explicativa de la heterocedasticidad). Supongamos que $z_{0.95}$ es el percentil 0.95 de la distribucion $\chi^2(1)$. Entonces, de acuerdo al resultado anterior:

$$Pr(LM > z_{0.95}) = 0.050$$

asintoticamente, bajo H_0.

En primer lugar, para que parametros esta magnitud es asintoticamente pivotal y para cuales no.

- H_0 fija $\alpha_2 = 0$
- α_1, Z, β, σ^2 y la distribucion de X pueden ser elegidos libremente. Porque?.
- La distribucion normal es necesaria

No sabemos mucho acerca de la pivotalidad en muestra finita.

Experimento:

1. Fijamos T (tamaño de muestra) y J (cantidad de replicaciones de montecarlo)
2. Computar J muestras de tamaño T de un modelo $Y = 1 + X + u$ bajo la hipotesis de homocedasticidad. Tomamos $\sigma^2 = 1$, u normales, X uniformes.
3. En cada paso computamos LM_{BP}
4. Finalmente, observamos cual es la proporcion de casos que rechazan la hipotesis nula usando $z_{0.95}$

Sesgo en el estimador de la varianza

```
.summarize

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          s2 |      1000    1.001777    .3343313   .2927112   2.272603
      sigma2 |      1000    .9516883    .3176147   .2780756   2.158973
```

```
sum, detail bpag
-------------------------------------------------------------
      Percentiles      Smallest

50%      .4260018                      Mean           .9466999
                        Largest        Std. Dev.      1.412939
75%      1.190816      12.43278
90%      2.580779      13.16408        Variance       1.996395
95%      3.585962      13.76204        Skewness       3.362467
99%      6.566663      15.51173        Kurtosis        20.6623

Percentiles chi-sq(1)
> qchisq(0.9,1)
[1] 2.705543
> qchisq(0.95,1)
[1] 3.841459
> qchisq(0.99,1)
[1] 6.634897
```