Pro Mathematica Vol. XVII, N° 34, 2003

MÉTODO DE STEIN

Elizabeth ***Doig***

Resumen

Se da una apreciación global del método de Stein[20] para la aproximación normal. Se empieza con los resultados básicos de las ecuaciones de Stein y luego se demuestran algunos teoremas que ilustran su utilidad

Sección Matemáticas, Departamento de Ciencias, PUCP.

Introducción

El método de Stein se describe en [20] para demostrar las aproximaciones normales. Este método evita las funciones características, pero en cambio se basa en una ecuación determinada para la distribución normal. La distribución de cualquier variable aleatoria sería entonces casi normal si satisface aproximadamente dicha ecuación.

1 El Método de Stein para las Aproximaciones Normales

Ilustraremos este método primero para el ejemplo original, de las aproximaciones normales. Descrito por Stein[20], con más detalle se puede encontrar en [21]; mejorándose los límites en Baldi[3]. Se puede resumir como sigue:

1. Una variable aleatoria I tiene distribución normal estándar; es decir, $N(0,1)$, si y sólo si para toda función f continua, se tiene
$$E[f'(I) - If(I)] = 0.$$
2. Sea I una variable aleatoria con Distribución Normal Estándar $N(0,1)$. Para cualquier función h continua, existe una función $f = f_h$ solución de la siguiente ecuación:
$$h(x) - E[h(I)] = f'(x) - xf(x) \text{ (Ecuación de Stein)}$$
tal que
$$||f|| \leq \sqrt{\tfrac{\pi}{2}}\, ||h - E[h(I)]||$$
$$||f'|| \leq (\sup h - \inf h)$$
$$||f''|| \leq 2||h'||.$$
En donde, $||f|| = \sup_{x\in\Re} |f(x)|$ denota la norma suprema.
3. Así que para cualquier variable aleatoria W, y cualquier función h continua, sustituyendo a x por W en la ecuación de Stein y tomando la esperanza en ambos lados se tiene
$$E[h(W)] - E[h(I)] = E[f'(W)] - E[Wf(W)]. \tag{2}$$

Ahora si H es una clase de convergencia determinante para la convergencia débil tal que, para toda $h \in H$ tenemos que $f_h \in H$, entonces tomando valor absoluto en la Ecuación (2) da

$$\sup_{h \in H} |E[h(W)] - E[h(I)]| = \sup_{f \in H} |E[f'(W)] - E[Wf(W)]|$$

y si el lado derecho de la ecuación, para alguna cantidad $W = W_n$, tiende a cero cuando $n \to \infty$, entonces obtenemos el teorema del límite central. En particular, podemos escoger a H como el espacio de todas las funciones continuas, acotadas, con continuidad a trozos o por partes, con las primeras derivadas acotadas.

Para ver porque este acercamiento podría ser útil, considere un ejemplo clásico. Sean las variables aleatorias $X, X_1, \ldots X_n$ i.i.d. con $E[X] = 0$ y $Var[X] = \frac{1}{n}$. Entonces

$$W = \sum_{k=1}^{n} X_k$$

tiene media cero y varianza 1. Poner:

$$W_k = W - X_k = \sum_{j \neq k} X_j.$$

Para evaluar el lado derecho de la ecuación (2) calculamos primero

$$E[Wf(W)] = \sum_{k=1}^{n} E[X_k f(W)]$$

$$= \sum_{k=1}^{n} E[X_k f(W_k)] + \sum_{k=1}^{n} E[X_k^2 \, f'(W_k)] + R$$

$$= \frac{1}{n} \sum_{k=1}^{n} E[f'(W_k)] + R,$$

donde usamos la expansión de Taylor, y R es el término del resto de Taylor tenemos

$$|R| \leq \|f''\| \sum_{k=1}^{n} E[|X_k|].$$

Así

$$E[f'(W)] - E[Wf(W)] = \frac{1}{n}\sum_{k=1}^{n} E[f'(W) - f'(W_k)] + R.$$

Aplicando otra vez la expansión de Taylor obtenemos el siguiente resultado [21]:

Teorema 1. *Sea I una variable aleatoria cuya distribución es la Distribución Normal Estándar $N(0,1)$. Para cualquier función continua h:*

$$|E[h(W)] - E[h(I)]| \leq \|h'\| \left(\frac{2}{\sqrt{n}} + \sum_{k=1}^{n} E\left[|X_k^3|\right] \right).$$

Notar que la cota en el lado derecho de la desigualdad no involucra ningún resultado asintótico; más bien es válido para cualquier número n. Así, aún cuando la convergencia no se pudiese mantener, aún se podría obtener una cota para la distancia.

El método de Stein ha sido particularmente útil para demostrar los resultados para las sumas de observaciones dependientes. Considere por ejemplo, el caso en que las variables aleatorias $X, X_1, \ldots, X_n$ con $E[X] = 0$ y $Var[X] = \frac{1}{n}$ tal que para cada X_k hay un conjunto S_k fijo tal que X_k es independiente de $\sigma(X_j, j \notin S_k)$. (Un caso especial sería el de m variables aleatorias dependientes.) Por simplicidad asumimos que $|S_k| = \gamma$ para algún γ, y que $X_k \leq \frac{C}{\sqrt{n}}$ para alguna constante C que no depende de k. Sea entonces la suma de n variables:

$$W = \sum_{k=1}^{n} X_k,$$

si γ es pequeño, entonces los sumandos son aproximadamente independientes, así que una aproximación normal se debería mantener. Demostraremos el siguiente resultado:

Teorema 2. *Sean las variables aleatorias $X, X_1, \ldots, X_n$ con $E[X] = 0$ y $Var[X] = \frac{1}{n}$, tal que para cada X_i hay un conjunto S_i tal que X_i es independiente de $\sigma(X_j, j \notin S_i)$. Asuma que $|S_i| = \gamma$ para algún γ, y* *que* $X_i \leq \frac{C}{\sqrt{n}}$ *para alguna constante* C *que no depende de* i. *Sea* I *una variable aleatoria con distribución Normal Estándar,* $N(0,1)$.

Para todas las funciones continuas, y acotadas h *con continuidad por partes y las primeras derivadas acotadas,*

$$|E[h(W)] - E[h(I)]| \leq ||h'|| \frac{10\gamma^2 C^3}{\sqrt{n}} + (\sup h - \inf h) \sum_{i=1}^{n} \sum_{j \in S_i, j \neq i} E[|X_i X_j|]$$

Notar que el Teorema 2 da una cota explícita en términos del tamaño de la vecindad y del número de observaciones, así como de las correlaciones. Si el valor de la covarianza es grande, uno aproximaría la suma más bien con una distribución normal que tiene como varianza $Var\left(\sum_{i=1}^{n} X_i\right)$; entonces el segundo término del error en el Teorema 2 desaparece.

Prueba del Teorema 2

Sea

$$W_i = \sum_{j \notin S_i} X_j \quad \text{y} \quad W_{i,j} = \sum_{k \notin S_i \cup S_j} X_k$$

Entonces W_i es independiente de X_i y $W_{i,j}$ es independiente de X_i y X_j. Similarmente a los cálculos anteriores, expandimos lo que está al lado derecho de la ecuación (2). Tenemos

$$\begin{aligned}
E[Wf(W)] &= \sum_{i=1}^{n} E[X_i f(W)] \\
&= \sum_{i=1}^{n} E[X_i f(W_i)] + \sum_{i=1}^{n} \sum_{j \in S_i} E[X_i X_j f'(W_i)] + R_1 \\
&= \sum_{i=1}^{n} \sum_{j \in S_i} E[X_i X_j f'(W_i)] + R_1
\end{aligned}$$

donde

$$\begin{aligned}
|R_1| &\leq ||f''|| \sum_{i=1}^{n} \sum_{j \in S_i} \sum_{k \in S_i} E[|X_i X_j X_k|] \\
&\leq ||f''|| \frac{\gamma^2 C^3}{\sqrt{n}}.
\end{aligned}$$

Es más,

$$\sum_{i=1}^{n} \sum_{j \in S_i} E[X_i X_j f'(W_i)] = \sum_{i=1}^{n} \sum_{j \in S_i} E[X_i X_j f'(W_{i,j})] + R_2,$$

donde

$$\begin{aligned} |R_2| &\leq \|f''\| \sum_{i=1}^{n} \sum_{j \in S_i} \sum_{k \in S_i \cup S_j} E[|X_i X_j|] E[X_k] \\ &\leq \|f''\| \frac{2\gamma^2 C^3}{\sqrt{n}} \end{aligned}$$

Usando el desarrollo de Taylor de nuevo, se obtendrá:

$$\begin{aligned} \sum_{i=1}^{n} \sum_{j \in S_i} E[X_i X_j f'(W_{i,j})] &= \sum_{i=1}^{n} \sum_{j \in S_i} E[X_i X_j] E[f'(W_{i,j})] \\ &= \sum_{i=1}^{n} \sum_{j \in S_i} E[X_i X_j] E[f'(W)] + R_3 \end{aligned}$$

con

$$|R_3| \leq \|f''\| \frac{2\gamma^2 C^3}{\sqrt{n}};$$

por lo tanto,

$$E[Wf(W)] = E[f'(W)] + \sum_{i=1}^{n} \sum_{j \in S_i j \neq i} E[X_i X_j] E[f'(W)] + R_1 + R_2 + R_3$$

Un estudio más detallado del método de Stein para las aproximaciones normales está en [18].

2 Método General de Stein

El método general de Stein se puede describir de la siguiente manera: Consiste en encontrar una buena caracterización (es decir, un operador A) de la distribución designada F tal que:

$$L(X) = F \Leftrightarrow E[Af(X)] = 0$$

para toda función f continua, donde X es una variable aleatoria, y L denota la ley de distribución de X. Consideremos la ecuación de Stein:

$$h(x) - E[h(X)] = Af(x). \tag{3}$$

Para cada función h lisa, encuentra una solución f correspondiente a esta ecuación. Entonces, para cualquier elemento aleatorio W, obtenemos

$$E[h(W)] - E[h(X)] = E[Af(W)].$$

Por lo tanto, para estimar la proximidad de W y X, es suficiente estimar $E[Af(W)]$ para toda solución posible f de (3). Aún más, acotando $E[Af(W)]$ para toda función f lisa automáticamente proporciona una cota para la distancia de $L(W)$ a F en una métrica lisa.

Para F la distribución normal estándar, el operador correspondiente es

$$Af(x) = f'(x) - xf(x).$$

Por supuesto que el operador también se podría definir como un operador de segundo orden, a saber $Af(x) = f''(x) - xf'(x)$, llamado el generador de Ornstein-Uhlenbeck. Barbour [6] sugiere emplear como operador A en la ecuación (3) al generador de un proceso de Markov, como esto proporciona una manera de buscar las soluciones de la ecuación (3). A partir de ahora se llamará el método generador. Suponga que se puede encontrar un proceso de Markov $(X(t))_{(t\geq 0)}$ con el generador A y la única distribución estacionaria F, tal que

$$L(X(t)) \xrightarrow{\omega} F(t \to \infty);$$

en donde $\xrightarrow{\omega}$ denota convergencia débil. Entonces, si una variable aleatoria X tiene distribución F,

$$E[Af(X)] = 0$$

para toda función f tal que Af está definida. Un método para resolver la ecuación (3) es proporcionada por Ethier y Kurtz [12]. Sea $(T_t)_{(t\geq 0)}$ el semigrupo de transición del Proceso de Markov $(X(t))_{(t\geq 0)}$. Entonces

$$T_t h - h = A\left(\int_0^t T_u h du\right).$$

Como $(T_t)_{(t\geq 0)}$ es un semigrupo de contracción fuertemente continuo, A está cerrado [12] y podríamos tomar los límites formalmente:

$$h(x) - E[h(X)] = -A\left(\int_0^\infty T_u h du\right).$$

Por lo tanto $f = -\int_0^\infty T_u h du$ sería una solución de (3), si esta expresión existe y sí $f \in D(A)$. En general este es sólo el caso para una cierta clase de funciones h.

Observaciones

Como se observa por lo expuesto anteriormente, el método de Stein es particularmente útil para demostrar los resultados para las sumas de observaciones dependientes. Además, así como se aplicó a la distribución Normal, también se ha generalizado a muchas otras distribuciones, como a la distribución de Poisson (Chen [10], Arratia [2], Barbour [8], Aldous [1]). Otras distribuciones incluyen la distribución uniforme (Diaconis [11]), la distribución binomial (Ehm [13]), la distribución de Poisson compuesta (Barbour [7], Barbour y Utev [9], Roos [19]), la distribución multinomial (Loh [14]), la distribución gamma (Luk [15]; para la distribución Ji- cuadrado (Mann [16])), y la distribución geométrica (Peköz [17]).

Una de las mayores ventajas del método de Stein es que mantiene las cotas inmediatas.

Referencias

[1] ALDOUS, D. (1989). *Stein's method in a two - dimensional coverage problem.* Statist. Probab. Lett. 8, 307 - 314.

[2] ARRATIA, R., GOlDSTEIN, L., AND GORDON, L. (1989). *Two moments suffice for Poisson approximations: the Chen - Stein method.* Ann. Probab. 17, 9-25.

[3] BALDI, P., RINOTT, Y., AND STEIN, C. (1989). *A normal approximation for the number of local maximal of a random function on a graph.* In Probability, Statistics and Mathematics, Papers in Honor of Samuel.

[4] BARBOUR, A.D. (1974). *On a functional central limit theorem for Markov population processes.* Adv. Appl. Probab. 6, 21 - 39.

[5] BARBOUR, A.D. (1975). *The duration of the closed stochastic epidemic.* Biometrika 62, 477-482.

[6] BARBOUR, A.D. (1990). *Stein's method for diffusion approximations.* Probab. Theory Relat. Fields 84, 297-322.

[7] BARBOUR, A.D., CHEN, L.H.Y., AND LOH, W. - L. (1992a). *Compound Poisson approximation for nonnegative random variables via Stein's method.* Ann. Probab. 20, 1843-1866.

[8] BARBOUR, A.D., HOLST, L., AND JANSON, S. (1992b). *Poisson Approximation.* Oxford Science Publications.

[9] BARBOUR, A.D., AND UTEV, S. (1998). *Solving the Stein equation in compound Poisson approximation.* Adv. Appl. Prob. 30, 449 - 475.

[10] CHEN, L.H.Y. (1975). *Poisson approximation for dependent trials.* Ann. Probab. 3, 534-545.

[11] DIACONIS, P. (1989). *An example for Stein's method.* Technical Report, Stanford University. Stat. Dept.

[12] ETHIER, S.N. AND KURTZ, T.G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

[13] EHM, W. (1991). *Binomial approximation to the Poisson binomial distribution.* Statist. Probab. Lett. 11, 7-16.

[14] LOH, W. - L. (1992).*Stein's method and multinomial approximation.* Ann. Appl. Probab. 2, 536 - 554.

[15] LUK, H.M. (1994). *Stein's method for the gamma distribution and related statistical applications.* Ph.D. thesis. University of Southern California, Los Angeles.

[16] MANN, B. (1997). *Stein's method for of a multinomial.* Preprint, Harvard.

[17] Pekoz, E. (1995). *Stein's method for geometric approximation.* Technical Report - 123. Stat. Dept. University of California, Berkeley.

[18] Reinert, G. (1998a). *Couplings for normal approximations with Stein's method.* In DIMACS Series in Discrete Mathematics and Theoretical Computer Science 41. 193-207.

[19] Roos, M. (1993). *Stein - Chen method for compound Poisson approximation.* Ph.D. thesis, University of Zurich.

[20] Stein, C. (1972). *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables.* Proc. Sixth Berkeley Symp. Math. Statist. Probab. 2, 583-602. Univ. of California Press, Berkeley.

[21] Stein, C. (1986). *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes - Monograph Series, Vol. 7, Hayward, California.

Elizabeth Doig
Sección Matemáticas
Departamento de Ciencias,
Pontificia Universidad Católica del Perú
edoig@pucp.edu.pe