Método del muestreo por importancia

puedo trabajar en más de una variable

la inicial

Sean $\boldsymbol{X}$ un vector aleatorio con función de densidad *nominal* f_1 y $g\colon \mathbb{R}^d \to \mathbb{R}$ una función medible. Entonces

$$\mu = \mathbb{E}_{f_1}[g(\boldsymbol{X})] = \int_{\mathbb{R}^d} g(\boldsymbol{x}) f_1(\boldsymbol{x})\, d\boldsymbol{x}$$

la nueva

El método de muestreo por importancia consiste en considerar una función de densidad *instrumental* f_2 para $\boldsymbol{X}$, de tal manera que f_1 sea absolutamente continua con respecto a f_2 (es decir, $f_2(\boldsymbol{x}) \neq 0$ para todo $\boldsymbol{x}$ tal que $f_1(\boldsymbol{x}) \neq 0$). Si denotamos el soporte de una función f como $\operatorname{sop}_f = \{\boldsymbol{x} \in \mathbb{R}^d \mid f(\boldsymbol{x}) \neq 0\}$ y definimos $L(\boldsymbol{x}) = \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})}$ como la *razón de verosimilitud*, se tiene que

para no dividir por 0

densidad nominal / densidad instrumental

$$\begin{aligned}
\mathbb{E}_{f_2}[g(\boldsymbol{x})L(\boldsymbol{x})] &= \int_{\mathbb{R}^d} g(\boldsymbol{x})L(\boldsymbol{x})f_2(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{\operatorname{sop}_{f_2}} g(\boldsymbol{x})L(\boldsymbol{x})f_2(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{\operatorname{sop}_{f_2}} g(\boldsymbol{x})f_1(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{\operatorname{sop}_{f_1}} g(\boldsymbol{x})f_1(\boldsymbol{x})\, d\boldsymbol{x} + \int_{\operatorname{sop}_{f_2}\setminus \operatorname{sop}_{f_1}} g(\boldsymbol{x})f_1(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{\operatorname{sop}_{f_1}} g(\boldsymbol{x})f_1(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} g(\boldsymbol{x})f_1(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \mathbb{E}[g(\boldsymbol{x})] \\
&= \mu
\end{aligned}$$

El soporte de f1 está incluido en el soporte de f2

soportes

donde f_2 vale distinto de 0

Se tiene entonces que el estimador

Es el método montecarlo aplicado a esa esperanza

$$\hat{\mu}_{\mathrm{is},f_2,n}(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{x}_i)L(\boldsymbol{x}_i)$$

is important sampling

donde $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{\text{i. i. d.}}{\sim} f_2$, es un estimador insesgado de μ.

Además,

$$\begin{aligned}\mathrm{Var}_{f_2}\big(\hat{\mu}_{\mathrm{is},f_2,n}(\boldsymbol{X})\big) &= \frac{1}{n}\mathrm{Var}_{f_2}(g(\boldsymbol{X})L(\boldsymbol{X}))\\ &= \frac{1}{n}\Big(\mathbb{E}_{f_2}[g(\boldsymbol{X})^2L(\boldsymbol{X})^2] - \mathbb{E}_{f_2}[g(\boldsymbol{X})L(\boldsymbol{X})]^2\Big)\\ &= \frac{1}{n}\left(\int_{\mathbb{R}^d} g(\boldsymbol{x})^2L(\boldsymbol{x})^2 f_2(\boldsymbol{x})\,d\boldsymbol{x} - \mu^2\right)\\ &= \frac{1}{n}\left(\int_{\mathbb{R}^d} g(\boldsymbol{x})^2L(\boldsymbol{x}) f_1(\boldsymbol{x})\,d\boldsymbol{x} - \mu^2\right)\\ &= \frac{1}{n}\big(\mathbb{E}_{f_1}[g(\boldsymbol{X})^2L(\boldsymbol{X})] - \mu^2\big)\end{aligned}$$

es le media de eso, definición de varianza

Como $\mathrm{Var}_{f_1}(\hat{\mu}_n(\boldsymbol{X})) = \frac{1}{n}(\mathbb{E}_{f_1}[g(\boldsymbol{X})^2] - \mu^2)$, se tiene que $\mathrm{Var}_{f_2}\big(\hat{\mu}_{\mathrm{is},f_2,n}(\boldsymbol{X})\big) < \mathrm{Var}_{f_1}(\hat{\mu}_n(\boldsymbol{X}))$ si y solo si $\mathbb{E}_{f_1}\big[g(\boldsymbol{X})^2L(\boldsymbol{X})\big] < \mathbb{E}_{f_1}[g(\boldsymbol{X})^2]$.

no puedo estimar nada si la esperanza vale 0

para que exista la esperanza

Teorema 1 *Si* $0 < \int_{\mathbb{R}^d} |g(\boldsymbol{x})| f_1(\boldsymbol{x})\,d\boldsymbol{x} < +\infty$, *entonces la elección de* f_2 *que minimiza el valor de* $\mathrm{Var}_{f_2}\big(\hat{\mu}_{\mathrm{is},f_2,n}(\boldsymbol{X})\big)$ *es*

$$f_2^*(\boldsymbol{x}) = \frac{|g(\boldsymbol{x})| f_1(\boldsymbol{x})}{\int_{\mathbb{R}^d} |g(\boldsymbol{x})| f_1(\boldsymbol{x})\,d\boldsymbol{x}}$$

la f_2 que tengo que elegir para obtener la mínima varianza posible

y la mínima varianza es

$$\left(\int_{\mathbb{R}^d} |g(\boldsymbol{x})| f_1(\boldsymbol{x})\,d\boldsymbol{x}\right)^2 - \left(\int_{\mathbb{R}^d} g(\boldsymbol{x}) f_1(\boldsymbol{x})\,d\boldsymbol{x}\right)^2$$

mi f2 se tiene que parecer al cociente, el denominador es una constante normalizadora, que luego no tiene problema

Parecerse es: que donde esa expresión tome un valor alto es donde mi densidad instrumental tiene que tomar valores mayores

Este resultado de optimalidad es, en realidad, un resultado formal: por ejemplo, cuando g es siempre positiva o negativa se alcanzaría varianza 0, pero esta elección óptima requeriría conocer $\int_{\mathbb{R}^d} g(\boldsymbol{x}) f_1(\boldsymbol{x})\,d\boldsymbol{x}$, que es la integral que se quiere calcular.

no puedo aplicar el método en la práctica

No obstante, desde un punto de vista práctico el resultado sugiere probar con densidades instrumentales que sean lo más similares a $|g(\boldsymbol{x})| f_1(\boldsymbol{x})$ que sea posible. Es decir, las tuplas $\boldsymbol{x}$ que deberían generarse con mayor probabilidad (las más «importantes») deberían ser aquellas con valores altos de $|g(\boldsymbol{x})| f_1(\boldsymbol{x})$. También hay que tener en cuenta que la varianza del método es menor cuanto menor sea $\mathbb{E}_{f_1}\big[g(\boldsymbol{X})^2L(\boldsymbol{X})\big]$. Esto quiere decir que, puesto que el denominador de L es f_2, una densidad instrumental mal elegida puede dar lugar a una estimación de gran varianza, incluso infinita. Para evitarlo, los valores cercanos a cero de esta densidad instrumental deberían cancelarse con valores cercanos a cero de la función g o, en su defecto, la densidad instrumental f_2 debería acercarse a cero más lentamente que la densidad nominal f_1.