Simulación II

Tema: Resumen de Probabilidad

Asociado a cada evento A de un espacio muestral S hay un número $P(A)$, la probabilidad de ocurrencia del evento A, que satisface los siguientes axiomas:

1. $0 \leq P(A) \leq 1$.
2. $P(S) = 1$.
3. $P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A), \qquad n = 1, 2, ..., \infty$.

La probabilidad conjunta de los eventos A y B (la probabilidad de que sucedan simultáneamente) se denota como $P(AB)$. La probabilidad condicional del evento A dado el evento B se escribe $P(A|B)$, y se cumple la relación:

$$P(A|B) = \frac{P(AB)}{P(B)}. \tag{1}$$

Y cuando los eventos A y B son independientes:

$$P(AB) = P(A)P(B). \tag{2}$$

Definition 0.1 *El valor esperado, o media, de una variable aleatoria discreta X que asume alguno de los valores posibles $x_1, x_2, ..., x_n$, es*

$$\mathbb{E}[X] := \sum_i x_i P\{X = x_i\}. \tag{3}$$

Definition 0.2 *El valor esperado, o media, para una variable continua X que puede tomar todos los valores x en $(-\infty, \infty)$, con función densidad de probabilidad f, se define como:*

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f(x) dx. \tag{4}$$

Definition 0.3 *Para una variable aleatoria X con media μ se define la varianza como*

$$Var(X) := \mathbb{E}[(X - \mu)^2]. \tag{5}$$

Definition 0.4 *La covarianza de dos variables aleatorias, X de media μ_x y Y de media μ_y, se define como*

$$Cov(X, Y) := \mathbb{E}[(X - \mu_x)(Y - \mu_y)]. \tag{6}$$

Theorem 0.1 *El valor esperado y la varianza (Definición 0.3) tienen las siguientes propiedades*

1. $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.
2. $\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$.

3. $Var(X) = \mathbb{E}[X^2] - \mu^2$.
4. $Var(aX + b) = a^2 Var(X)$.

donde a y b son constantes y X_1 y X_2 son variables aleatorias.

Theorem 0.2 *La varianza y la covarianza (Definición 0.4) satisfacen las siguientes propiedades*

1. $Cov(X,Y) = \mathbb{E}[XY] + \mathbb{E}[X]\mathbb{E}[Y]$.
2. $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$.
3. $Var(X) \geq 0$.

Así, cuando X y Y son independientes, $Cov(X,Y) = 0$ y se tiene

$$Var(X + Y) = Var(X) + Var(Y).$$

Algunos de los resultados más importantes de la teoría de probabilidad son las desigualdades, pues con base en ellas se pueden hacer inferencias y determinar intervalos de confianza. Aquí nos interesan principalmente los resultados para variables continuas. Uno muy útil es el siguiente.

Theorem 0.3 ***Desigualdad de Markov.***
Si X toma solo valores no negativos, entonces para cualquier $a > 0$

$$P\{X \geq a\} \leq \frac{\mathbb{E}[X]}{a}. \tag{7}$$

Demostración. Si X es una variable aleatoria no negativa, entonces

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{\infty} x f(x) dx,$$

y separando en dos segmentos el intervalo de integración:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^a x f(x) dx + \int_a^{\infty} x f(x) dx \\
&\geq \int_a^{\infty} x f(x) dx \\
&\geq \int_a^{\infty} a f(x) dx, \ \ \text{pues} \ \ x f(x) \geq a f(x) \ \ \text{si} \ \ x \geq a \\
&= a \int_a^{\infty} f(x) dx = aP\{X \geq a\},
\end{aligned}$$

de donde se obtiene el resultado deseado. ■

La desigualdad de Markov permite probar fácilmente el siguiente resultado, de mayor utilidad práctica.

Theorem 0.4 ***Desigualdad de Chebyshev.***
Si X es una variable aleatoria con media μ y varianza σ^2, entonces para cualquier $k > 0$

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}. \tag{8}$$

Demostración. Se obtiene aplicando la desigualdad de Markov (ec. 7) a la variable no negativa $\frac{(X-\mu)^2}{\sigma^2}$, que tiene media igual a 1, como puede comprobarse a partir de la definición de la varianza y las propiedades del valor esperado (Teorema 0.1). ■

La desigualdad de Chebyshev (ecuación 8) permite, considerando la hipótesis que la varianza siempre tiene valores finitos, arribar a los resultados más importantes para el método de Monte Carlo ordinario: las leyes de los grandes números (Teoremas 0.5 y 0.6).

Theorem 0.5 ***Ley débil de los grandes números.***
Sea $X_1, X_2, ...$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con media μ. Entonces, para cada $\epsilon > 0$

$$P\Big\{\Big|\frac{X_1 + ... + X_n}{n} - \mu\Big| < \epsilon\Big\} \to 0, \quad \textit{cuando} \;\; n \to \infty. \tag{9}$$

La generalización de este resultado es:

Theorem 0.6 ***Ley fuerte de los grandes números.***
Bajo las condiciones del teorema anterior, se tiene con probabilidad 1, que

$$\lim_{n\to\infty} \frac{X_1 + ... + X_n}{n} = \mu. \tag{10}$$

Y un resultado de gran utilidad para realizar inferencia estadística es el:

Theorem 0.7 ***Teorema central del límite.***
Dada una sucesión de variables aleatorias independientes e idénticamente distribuidas, $X_1, X_2, ...$, con media finita μ y varianza finita σ^2 se cumple que

$$\lim_{n\to\infty} P\Big\{\frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}} < x\Big\} = \phi(x), \tag{11}$$

donde $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-x^2}{2}} dx$, para $-\infty < x < \infty$, es la distribución normal o gaussiana.

La ley fuerte de los grandes números es el fundamento de la integración por el método de Monte Carlo. Nótese, en primer lugar, que la ec. (4) permite interpretar una integral $\int_a^b g(x)dx$ como el valor esperado, μ, de la función $g(x)$ respecto a una distribución de probabilidad uniforme definida para $a \leq x \leq b$; es decir, si $f(x) = Unif(a, b) = \frac{1}{b-a}$, multiplicando la integral anterior por el factor $(b - a)$, se tiene

$$\int_a^b g(x)dx = (b-a)\int_a^b g(x)f(x)dx = (b-a)\mathbb{E}[g(X)] \;\text{ con }\; X \sim Unif(a, b).$$

Por otro lado, la ley fuerte de los grandes números (ec. 10), permite aproximar este último valor esperado mediante el promedio de una muestra de tamaño n de la variable aleatoria distribuida uniformemente, si se toma n suficientemente grande. Así:

$$\int_a^b g(x)dx = (b-a) \lim_{n\to\infty} \frac{X_1 + ... + X_n}{n} \text{ con } X \sim Unif(a,b).$$

La cuestión relevante, ahora, es cómo generar las variables aleatorias uniformemente distribuidas, $X_1, ..., X_n$. Exploramos esto en la siguiente sección.