

## 2.- Algunos conceptos básicos de probabilidad y estadística

### 2.1- ¿Para qué?

Como hemos visto en la sección precedente, la probabilidad y la estadística son necesarias para:

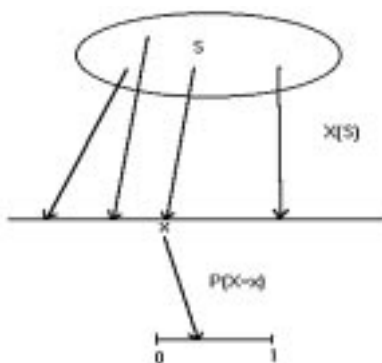
- determinar las distribuciones de probabilidad de entrada
- generar variables aleatorias de acuerdo con estas distribuciones
- realizar la validación del modelo
- analizar estadísticamente los datos de salida de la simulación
- diseñar los experimentos a simular.

En esta sección introduciremos algunos conceptos básicos de probabilidad y estadística relevantes a la hora de diseñar, realizar y analizar los resultados de una simulación.

Para ello, adoptaremos el convenio de designar con letras mayúsculas a las variables aleatorias y con minúsculas sus valores.

### 2.2- Espacio muestral y variable aleatoria

Se define el *espacio muestral*  $S$  de un experimento aleatorio como el conjunto de todos los resultados posibles del experimento. Los elementos del espacio deben excluirse mutuamente, es decir, sólo uno de ellos puede ser resultado del experimento cada vez.



Una *variable aleatoria*  $X(S)$  es una función que asocia un número a cada uno de los elementos del espacio muestral.

Una *variable aleatoria* se dice *discreta* cuando su espacio muestral tiene un conjunto numerable de elementos, es decir, cuando sólo puede tomar un conjunto de valores finito o bien un conjunto de valores infinito numerable.

En caso contrario, se dice *continua*. Las variables aleatorias continuas son una idealización matemática, ya que cualquier procedimiento de medida que se emplee tendrá un límite de precisión. Igualmente, al programar cualquier algoritmo de generación de números aleatorios éstos se obtendrán con un determinado número finito de cifras decimales.

## 2.3- Probabilidad

Veamos tres definiciones de probabilidad: la definición clásica, la definición como frecuencia relativa y la definición axiomática.

La definición clásica de probabilidad es: Casos Favorables / Casos Posibles. Por ejemplo, en el lanzamiento de una moneda:

$$\begin{aligned}\text{Casos Posibles} &= \{\text{Cara, Cruz}\} \\ P\{\text{Cara}\} &= 1/2 \\ P\{\text{Cruz}\} &= 1/2\end{aligned}$$

En el caso del lanzamiento de un dado:

$$\begin{aligned}\text{Casos Posibles} &= \{1, 2, 3, 4, 5, 6\} \\ P\{1\} = P\{2\} = P\{3\} = P\{4\} = P\{5\} = P\{6\} &= 1/6 \\ P\{\text{Impar}\} &= 3/6 \\ P\{\text{Mayor que 1}\} &= 5/6\end{aligned}$$

Sin embargo, vemos que esta definición sólo es válida cuando el número de casos posibles es finito y además todos ellos son equiprobables. Así, por ejemplo, sería ilógico calcular la probabilidad de sufrir un accidente de avión de la forma:

$$\begin{aligned}\text{Casos Posibles} &= \{\text{Accidente, No accidente}\} \\ P\{\text{Accidente}\} &= 1/2\end{aligned}$$

Obsérvese que al afirmar que “la definición clásica de *probabilidad* es sólo aplicable cuando todos los eventos posibles son igualmente *probables*” estamos empleando en la definición el concepto definido. Sin embargo, la definición clásica fue introducida como consecuencia del *principio de razón insuficiente*: “si no poseemos ningún conocimiento a priori, debemos asumir que los posibles resultados del experimento son equiprobables”.

La definición de la probabilidad como frecuencia relativa es:  $P\{A\} = \lim_{n \rightarrow \infty} \frac{n_A}{n}$

donde  $n_A$  es el número de experimentos en que ocurre  $A$  y  $n$  es el número total de experimentos. Sin embargo, un experimento físico no puede repetirse infinitas veces, con lo cual, la probabilidad de un evento  $A$ ,  $P\{A\}$ , puede interpretarse:

Si el experimento se realiza  $n$  veces, de las cuales el evento  $A$  sucede  $n_A$  veces, entonces,

con un elevado grado de certeza, la frecuencia relativa  $\frac{n_A}{n}$  de ocurrencia del evento  $A$  esta

próxima a  $P\{A\}$ ,  $P\{A\} \approx \frac{n_A}{n}$ , supuesto que  $n$  es suficientemente grande.

Esta interpretación es imprecisa. Los términos *con un elevado grado de certeza, próxima y suficientemente grande* no tienen un significado claro. Sin embargo, no puede evitarse esta pérdida de precisión, con lo cual, si se usa  $P\{A\} = \lim_{n \rightarrow \infty} \frac{n_A}{n}$  para definir  $P\{A\}$ , el límite debe aceptarse como una hipótesis, no como un número que pueda ser determinado experimentalmente.

La definición axiomática de la probabilidad fue introducida en 1933 por Kolmogoroff y se basa únicamente en los siguientes tres postulados:

- 1.- La probabilidad  $P\{A\}$  de un evento  $A$  es un número positivo asignado a este evento:  $P\{A\} \geq 0$ .
- 2.- La probabilidad de un evento  $E$  que se obtiene siempre en todas las repeticiones del experimento es uno:  $P\{E\} = 1$ .
- 3.- La unión  $A + B$  de dos eventos  $A$  y  $B$  es el evento que ocurre cuando sucede  $A$  o  $B$  o suceden ambos. Se dice que los eventos  $A$  y  $B$  son mutuamente excluyentes si la ocurrencia de uno de ellos excluye la ocurrencia del otro.  
Si los eventos  $A$  y  $B$  son mutuamente excluyentes, entonces:  $P\{A + B\} = P\{A\} + P\{B\}$ .

Esta definición no tiene ambigüedades conceptuales y es aceptada generalmente como superior a las dos definiciones anteriores.

Notaremos:

$$P\{X = x_i\} = p_x(x_i) \quad \text{para } i: 1, 2, \dots$$

la probabilidad de que una variable aleatoria discreta  $X$  tome un determinado valor  $x_i$ , es decir, la probabilidad del evento  $\{X = x_i\}$ .

La probabilidad de que la variable aleatoria discreta  $X$  tome un valor menor o igual que  $x$  viene dado por su función de probabilidad acumulada  $F_X(x)$ , definida:

$$P\{X \leq x\} = \sum_{i: x_i \leq x} p_x(x_i) = F_X(x)$$

La probabilidad de que el valor de la variable aleatoria discreta  $X$  este contenido en el intervalo  $[a, b]$  puede calcularse de la forma:

$$P\{a \leq X \leq b\} = \sum_{i: a \leq x_i \leq b} p_x(x_i) = F_X(b) - F_X(a)$$

En el ejemplo del inventario, la cantidad de producto demandado es una variable aleatoria discreta,  $X$ , que puede tomar los valores  $X = \{1, 2, 3, 4\}$ , con probabilidad:

$$P\{X = 1\} = p_x(1) = \frac{1}{6}; \quad P\{X = 2\} = p_x(2) = \frac{2}{6};$$

$$P\{X = 3\} = p_x(3) = \frac{2}{6}; \quad P\{X = 4\} = p_x(4) = \frac{1}{6}$$

La función probabilidad acumulada es:

$$P\{X \leq x\} = F_x(x) = \begin{cases} 0 & \text{para } x < 1 \\ \frac{1}{6} & \text{para } 1 \leq x < 2 \\ \frac{3}{6} & \text{para } 2 \leq x < 3 \\ \frac{5}{6} & \text{para } 3 \leq x < 4 \\ \frac{6}{6} & \text{para } 4 \leq x \end{cases}$$

Cuando el espacio muestra consta de un número infinito e incontable de elementos, entonces sus probabilidades no pueden determinarse en términos de las probabilidades de los eventos individuales. Lo que se hace en este caso, en que la variable aleatoria es continua, es asignar probabilidad a los eventos  $\{X \leq x\}$ .

Se define la probabilidad del evento  $\{X \leq x_1\}$  como la integral:

$$P\{X \leq x_1\} = F_x(x_1) = \int_{-\infty}^{x_1} f_x(x) dx$$

donde  $f_x$  es una función no negativa,  $f_x(x) \geq 0$  para todo  $x$ , que satisface:  $\int_{-\infty}^{\infty} f_x(x) dx = 1$ .

La función  $F_x$  recibe el nombre de *distribución de probabilidad acumulada* y  $f_x$  el de densidad de probabilidad.

La probabilidad del evento  $\{a < X \leq b\}$  viene dada por:

$$P\{a < X \leq b\} = \int_a^b f_x(x) dx = F_x(b) - F_x(a)$$

Obsérvese que los eventos  $\{X \leq x_1\}$  y  $\{x_1 < X \leq x_2\}$  son mutuamente excluyentes y su unión es  $\{X \leq x_2\}$ , satisfaciéndose:

$$P\{X \leq x_1\} + P\{x_1 < X \leq x_2\} = P\{X \leq x_2\}$$

Como la función  $f_x$  esta acotada, la integral  $\int_{x_1}^{x_2} f_x(x) dx$  tiende a cero cuando  $x_1 \rightarrow x_2$ .

Esto conduce a la conclusión de que la probabilidad de un evento  $\{x_2\}$ , consistente en que obtener un determinado valor, es cero para cualquier valor de  $X$ . La probabilidad de los eventos elementales es cero, aunque la probabilidad de la unión de todos ellos es uno.

## 2.4- Probabilidad condicionada. Teorema de Bayes.

Supuesto que el evento  $E_1$  ha ocurrido, la probabilidad de que ocurra el evento  $E_2$  viene dada por la probabilidad condicional  $P\{E_2|E_1\}$ .

Si los dos eventos son estadísticamente independientes (la ocurrencia de un evento no afecta a la probabilidad de que suceda el otro), entonces:

$$P\{E_2|E_1\} = P\{E_2\}$$

Si no son estadísticamente independientes:

$$P\{E_2|E_1\} = \frac{P\{E_2E_1\}}{P\{E_1\}}$$

donde  $P\{E_2E_1\}$  nota la probabilidad de que se produzcan los dos eventos  $E_1$  y  $E_2$ .

Despejando de la expresión anterior obtenemos:

$$P\{E_1E_2\} = P\{E_1\}P\{E_2|E_1\}.$$

que, si  $E_1$  y  $E_2$  son independientes, es equivalente a:

$$P\{E_1E_2\} = P\{E_1\}P\{E_2\}$$

Análogamente, la probabilidad de que se produzcan los eventos  $E_1$ ,  $E_2$  y  $E_3$  es:

$$P\{E_1E_2E_3\} = P\{E_1\}P\{E_2E_3|E_1\} = P\{E_1\}P\{E_2|E_1\}P\{E_3|E_1E_2\}$$

que, si los eventos  $E_1$ ,  $E_2$  y  $E_3$  son independientes es equivalente a:

$$P\{E_1E_2E_3\} = P\{E_1\}P\{E_2\}P\{E_3\}$$

Dadas las variables aleatorias discretas  $X_1, X_2, \dots, X_n$  independientes, la probabilidad de que  $X_1$  valga  $x_1$  y de que  $X_2$  valga  $x_2$  y ... y de que  $X_n$  valga  $x_n$  viene dada por la *probabilidad conjunta*:

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = P\{X_1 = x_1\}P\{X_2 = x_2\} \dots P\{X_n = x_n\}$$

Análogamente, para variables aleatorias continuas  $X_1, X_2, \dots, X_n$  independientes, la *función densidad de probabilidad conjunta* es:

$$f_X(x_1, x_2, \dots, x_n) = f_X(x_1)f_X(x_2) \dots f_X(x_n)$$

y la *distribución de probabilidad acumulada* para los valores  $x_1, x_2, \dots, x_n$ , definida como:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\}$$

vale, para variables aleatorias continuas  $X_1, X_2, \dots, X_n$  independientes:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i) \quad \text{para cualesquiera } x_1, \dots, x_n$$

**Ejemplo.** Una caja contiene 2 bolas amarillas, 3 rojas y 5 negras. Los eventos  $E_1$ ,  $E_2$  y  $E_3$  representan el hecho de que se extraiga de la caja una bola amarilla, una roja o una negra respectivamente. Las bolas son extraídas de la caja una tras otra. Las bolas negras y rojas son introducidas de nuevo en la caja inmediatamente después de ser sacadas, mientras que las amarillas no son devueltas.

La probabilidad de que las dos primeras bolas sean una roja seguida de una negra es:

$$p(E_2 E_3) = p(E_2) p(E_3 | E_2) = p(E_2) p(E_3) = \frac{3}{10} \frac{5}{10}$$

La probabilidad de que las dos primeras bolas sean una amarilla seguida de una roja es:

$$p(E_1 E_2) = p(E_1) p(E_2 | E_1) = \frac{2}{10} \frac{3}{9}$$

La probabilidad de que las dos primeras bolas sean una roja seguida de una amarilla es:

$$p(E_2 E_1) = p(E_2) p(E_1 | E_2) = \frac{3}{10} \frac{2}{10}$$

*Teorema de Bayes:*

Sean  $E_1, E_2, \dots, E_k$   $k$  eventos diferentes, de modo que cualquiera de ellos produce la ocurrencia de otro evento  $A$ . Entonces:

$$P\{E_i | A\} = \frac{P\{E_i\} P\{A | E_i\}}{\sum_{j=1}^k P\{E_j\} P\{A | E_j\}} \quad \text{para } i: 1, 2, \dots, k$$

El teorema de Bayes es muy útil para el cálculo de la probabilidad condicionada del evento  $E_i$ , cuando el evento  $A$  se ha producido, a partir de la probabilidad incondicional de  $E_i$  y la probabilidad condicional de  $A$  supuesto que se ha producido cada uno de los  $E_i$ .

## Propiedades de la probabilidad

$$\begin{aligned}
0 &\leq P\{A\} \leq 1 \\
P\{E\} &= 1 \\
P\{A + B\} &= P\{A\} + P\{B\}, \text{ si } AB = \emptyset \\
P\{A + B\} &= P\{A\} + P\{B\} - P\{AB\}, \text{ si } AB \neq \emptyset \\
P\{A|B\} &= \frac{P\{AB\}}{P\{B\}} \\
\text{Si } A \text{ y } B \text{ son independientes: } &P\{AB\} = P\{A\}P\{B\}
\end{aligned}$$

## 2.5- Media, varianza y desviación típica

La *media, esperanza o valor esperado* de una variable aleatoria  $X$ , que notaremos  $E(X)$ , es una propiedad de la distribución de probabilidad y se define como:

$$E(X) = \begin{cases} \sum_{j=1}^{\infty} x_j p_X(x_j) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

**Ejemplo.** Si  $X$  toma valores  $\{1, 2, 3, 4, 5, 6\}$  con probabilidad  $1/6$ , entonces:

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Sea  $g$  una función real de variable real  $g: \Re \rightarrow \Re$  y sea  $X$  una variable aleatoria. El valor esperado de  $g(X)$  se define como:

$$E(g(X)) = \begin{cases} \sum_{j=1}^{\infty} g(x_j) p_X(x_j) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

Linealidad. Sea  $X$  una variable aleatoria y sean  $g_1, \dots, g_n$  funciones reales de variable real.

$$E(a_1 g_1(x) + a_2 g_2(x) + \dots + a_n g_n(x)) = a_1 E(g_1(x)) + a_2 E(g_2(x)) + \dots + a_n E(g_n(x))$$

La cantidad  $E(X^n)$ ,  $n \geq 1$ , se llama *momento de orden  $n$*  de  $X$  y se calcula:

$$E(X^n) = \begin{cases} \sum_{\text{todo } j} x_j^n p_X(x_j) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} x^n f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

La *varianza* de una variable aleatoria  $X$ , con valor esperado  $E(X) = \mu$ , discreta o continua,  $\sigma^2(X)$  o  $\text{Var}(X)$ , es una medida de la dispersión de la variable aleatoria respecto de su media.

Se define la varianza:

$$\begin{aligned}\text{Var}(X) &= \sigma^2(X) = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) = E(X^2) - 2\mu\mu + \mu^2 = E(X^2) - \mu^2\end{aligned}$$

es decir, se define como el valor esperado de la variable aleatoria cuadrado de  $X$ , menos el cuadrado del valor esperado de  $X$ . La varianza es siempre un número no negativo.

Dadas las variables aleatorias  $X_1, X_2, \dots, X_n$  independientes, y definida la función lineal  $\sum_{i=1}^n a_i X_i$ , donde  $a_i$  son constantes, se cumple:

$$\text{var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{var}(X_i)$$

En general, dadas dos variables aleatorias  $X, Y$  de medias  $E(X) = \mu_X$  y  $E(Y) = \mu_Y$ , la varianza de la suma  $X+Y$  vale:

$$\begin{aligned}\text{var}(X + Y) &= E[(X + Y - \mu_X - \mu_Y)^2] = \\ &= E[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)] = \\ &= E[(X - \mu_X)^2] + E[(Y - \mu_Y)^2] + 2E[(X - \mu_X)(Y - \mu_Y)] = \\ &= \text{var}(X) + \text{var}(Y) + 2\text{Cov}(X, Y)\end{aligned}$$

Donde se define la covarianza de las variables aleatorias  $X, Y$  de medias  $E(X) = \mu_X$  y  $E(Y) = \mu_Y$  de la forma:

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] = E[XY - \mu_X Y - X\mu_Y + \mu_X \mu_Y] = \\ &= E[XY] - \mu_X E[Y] - E[X]\mu_Y + \mu_X \mu_Y = E[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y = \\ &= E[XY] - \mu_X \mu_Y\end{aligned}$$

La covarianza  $\text{Cov}(X, Y)$  de dos variables aleatorias independientes  $X, Y$  vale cero, con lo cual:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

Más adelante volveremos sobre el tema de la covarianza.

La *desviación típica o estandar*,  $\sigma(X)$ , se define como la raíz cuadrada de la varianza.



## 2.6- Estimador de la media y de la varianza

Supongamos que  $X_1, X_2, \dots, X_n$  son muestras aleatorias idénticamente distribuidas con media  $\mu$  y varianza  $\sigma^2$ .

La *media aritmética* de  $n$  muestras aleatorias  $X_1, X_2, \dots, X_n$  es también una variable aleatoria, definida como:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

El valor esperado de la distribución de la media aritmética es igual a la media  $\mu$  (no depende del número de muestras,  $n$ ):

$$E(\bar{X}) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{E(X_1) + \dots + E(X_n)}{n} = \frac{n\mu}{n} = \mu$$

Se dice, por ello, que la media aritmética de las muestras aleatorias idénticamente distribuidas  $X_1, X_2, \dots, X_n$  (no forzosamente independientes) es un estimador no sesgado de la media de la distribución de las muestras,  $\mu$ . En general, se dice que el estimador de una magnitud (media, varianza, etc.) es no sesgado cuando el valor esperado del estimador coincide con la magnitud.

Para conocer la dispersión de la variable aleatoria media aritmética,  $\bar{X}$ , respecto de su valor esperado,  $\mu$ , debemos calcular la varianza de la distribución de la media aritmética:

$$\begin{aligned} \text{var}(\bar{X}) &= E\left[\left(\frac{X_1 + \dots + X_n}{n} - \mu\right)^2\right] = E\left[\left(\frac{X_1 + \dots + X_n - n\mu}{n}\right)^2\right] = \\ &= E\left[\left(\frac{(X_1 - \mu) + \dots + (X_n - \mu)}{n}\right)^2\right] = \frac{1}{n^2} E\left[\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i \neq j} (X_i - \mu)(X_j - \mu)\right] = \\ &= \frac{1}{n^2} \sum_{i=1}^n E[(X_i - \mu)^2] + \frac{1}{n^2} \sum_{i \neq j} E[(X_i - \mu)(X_j - \mu)] = (*) \\ &= \frac{1}{n^2} \sum_{i=1}^n E[(X_i - \mu)^2] + \frac{1}{n^2} \sum_{i \neq j} \underbrace{E[(X_i - \mu)]}_{=0} \underbrace{E[(X_j - \mu)]}_{=0} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

(\*):  $X_1, X_2, \dots, X_n$  muestras aleatorias independientes de una misma distribución de media  $\mu$  y varianza  $\sigma^2$ .

La varianza de la distribución de la variable aleatoria  $\bar{X}$  es inversamente proporcional al número de muestras: al aumentar el número de las muestras,  $n$ , su media aritmética,  $\bar{X}$ , va desviándose cada vez menos de la media de la distribución de probabilidad de las muestras,  $\mu$ .

Dado que  $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ , para obtener una estimación de  $\text{Var}(\bar{X})$  debemos primeramente estimar la varianza de la distribución de probabilidad de las muestras,  $\sigma^2$ .

Veamos que la varianza de las muestras aleatorias  $X_1, X_2, \dots, X_n$ ,

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

es un estimador sesgado, y por tanto no válido, de la varianza de la distribución de las muestras,  $\sigma^2$ . Para ello, demostraremos que el valor esperado de  $S^2$ ,  $E(S^2)$ , no es igual a  $\sigma^2$ .

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = E\left(\frac{1}{n} \sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2\right) = \\ &= E\left(\frac{1}{n} \sum_{i=1}^n [(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)]\right) = \\ &= \frac{1}{n} E\left(\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2 \sum_{i=1}^n (X_i - \mu)(\bar{X} - \mu)\right) \end{aligned}$$

operando por separado el último sumatorio:

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)(\bar{X} - \mu) &= \sum_{i=1}^n X_i \bar{X} - \sum_{i=1}^n X_i \mu - \sum_{i=1}^n \mu \bar{X} + \sum_{i=1}^n \mu^2 = \\ &= n\bar{X}^2 - n\bar{X}\mu - n\mu\bar{X} + n\mu^2 = n(\bar{X}^2 - 2\bar{X}\mu + \mu^2) = n(\bar{X} - \mu)^2 = \\ &= \sum_{i=1}^n (\bar{X} - \mu)^2 \end{aligned}$$

sustituyendo obtenemos:

$$\begin{aligned} E(S^2) &= \frac{1}{n} E\left(\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2 \sum_{i=1}^n (\bar{X} - \mu)^2\right) = \\ &= \frac{1}{n} E\left(\sum_{i=1}^n (X_i - \mu)^2 - \sum_{i=1}^n (\bar{X} - \mu)^2\right) = \frac{1}{n} \sum_{i=1}^n (E[(X_i - \mu)^2] - E[(\bar{X} - \mu)^2]) = \\ &= \frac{1}{n} \sum_{i=1}^n (\sigma^2 - \sigma_{\bar{X}}^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \end{aligned}$$

Como vemos,  $S^2$  es una estimación sesgada de  $\sigma^2$ , lo que significa que si tomamos repetidas muestras de tamaño  $n$  y se promedian las varianzas muestrales resultantes, el promedio no se aproximará en valor a la verdadera varianza, sino que, de modo persistente, será demasiado pequeño debido al factor  $\frac{n-1}{n}$ . Este factor adquiere importancia cuando el tamaño de la muestra,  $n$ , es pequeño.

Para eliminar el sesgo en  $S^2$  basta con multiplicar  $S^2$  por  $\frac{n}{n-1}$ , y la cantidad que resulte utilizarla como estimación de  $\sigma^2$ :  $E\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}E(S^2) = \sigma^2$

Se define la *cuasivarianza muestral* de las muestras aleatorias  $X_1, X_2, \dots, X_n$ , que notaremos  $s^2$ , como:

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

y, en el caso en que las variables aleatorias  $X_1, X_2, \dots, X_n$  sean IID, es un estimador no sesgado de la varianza de la distribución de las muestras,  $\sigma^2$ :  $E(s^2) = \sigma^2$ .

Volviendo al problema de estimar la varianza de la distribución de  $\bar{X}$ ,  $Var(\bar{X}) = \frac{\sigma^2}{n}$ , podemos sustituir  $\sigma^2$  por su estimador,  $s^2$ . Obtenemos:

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

que constituye una medida del “agrupamiento”, de los distintos valores de  $\bar{X}$  que obtendríamos para diferentes muestras, entorno a la media de la distribución de las muestras,  $\mu$ .

Sean  $X_1, \dots, X_n$  variables aleatorias IID con media  $\mu$  finita. Entonces, la ley de los grandes números establece que:

$$\bar{X} \xrightarrow[n \rightarrow \infty]{} \mu \quad \text{con probabilidad 1}$$

Sin embargo, *los datos de salida de una simulación generalmente no son independientes*, con lo cual la discusión anterior sobre la estimación de la varianza de la media muestral, válida para observaciones IID, no lo será en este caso. Sin embargo, veremos que es posible agrupar los datos de salida en nuevas “observaciones” a las cuales si pueden aplicarse las fórmulas de la estadística clásica, basadas en observaciones IID.

También veremos cómo establecer intervalos de confianza para la media y la varianza. Es decir, cómo, a partir de las medidas experimentales obtener los intervalos que contienen a la media y a la varianza respectivamente con una probabilidad dada.

## 2.7- Covarianza y correlación

La *covarianza* entre dos variables aleatorias  $X_i, X_j$  ( $i:1,2,\dots,n; j:1,2,\dots,n$ ) es una medida de su dependencia lineal. Se define:

$$C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_i X_j) - \mu_i \mu_j$$

Obsérvese que las covarianzas son simétricas,  $C_{ij} = C_{ji}$ , y que si  $i=j$ , entonces,  $C_{ii} = \sigma_i^2$ .

Cuando la covarianza de dos variables aleatorias  $X_i, X_j$  es cero,  $C_{ij} = 0$ , se dice que son variables aleatorias no correlacionadas. *Las variables aleatorias independientes siempre están no correlacionadas*. Lo inverso, en general, no es cierto, excepto si sus distribuciones son normales.

Cuando  $C_{ij} > 0$ , se dice que  $X_i$  y  $X_j$  están *correlacionadas positivamente*. En este caso, como  $C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$ , el hecho de que  $C_{ij} > 0$  significa que tiene tendencia a ocurrir  $\begin{cases} X_i > \mu_i \\ X_j > \mu_j \end{cases}$  y  $\begin{cases} X_i < \mu_i \\ X_j < \mu_j \end{cases}$ . Con lo cual, para variables aleatorias correlacionadas positivamente, si una es grande, la otra también tiene tendencia a ser grande.

Cuando  $C_{ij} < 0$ , se dice que  $X_i$  y  $X_j$  están *correlacionadas negativamente*. Lo cual significa que tiene tendencia a ocurrir  $\begin{cases} X_i > \mu_i \\ X_j < \mu_j \end{cases}$  y  $\begin{cases} X_i < \mu_i \\ X_j > \mu_j \end{cases}$ . Con lo cual, para variables aleatorias correlacionadas negativamente, si una es grande, la otra tiene tendencia a ser pequeña.

Si  $X_1, \dots, X_n$  son variables aleatorias de salida de la simulación (por ejemplo,  $X_i$  puede ser el tiempo de espera  $D_i$  del cliente  $i$ -ésimo en la cola), a menudo interesa no sólo conocer las medias  $\mu_i$  y las varianzas  $\sigma_i^2$ , para  $i:1,2,\dots$ , sino también una medida de la dependencia de  $X_i$  y  $X_j$  para  $i \neq j$ .

Si bien la covarianza da una medida de la dependencia entre las variables aleatorias, es, a su vez, dependiente de las unidades en que éstas estén expresadas, lo que hace confusa su interpretación. Debido a esto, se define la *correlación* entre las variables aleatorias  $X_i$  y  $X_j$ ,  $\rho_{ij}$ , como:

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}} \quad \text{para } \begin{cases} i:1,2,\dots,n \\ j:1,2,\dots,n \end{cases}$$

como una medida de la dependencia entre  $X_i$  y  $X_j$ . Como  $\sqrt{\sigma_i^2 \sigma_j^2}$  es positivo,  $\rho_{ij}$  y  $C_{ij}$  tienen el mismo signo. Puede demostrarse que  $-1 \leq \rho_{ij} \leq 1$  para todo  $i, j$ . Si  $\rho_{ij}$  está próximo a  $+1$ , entonces  $X_i$  y  $X_j$  están fuertemente correlacionadas positivamente. Por el contrario, si  $\rho_{ij}$  está próximo a  $-1$ , entonces  $X_i$  y  $X_j$  están fuertemente correlacionadas negativamente.

## 2.8- Procesos estocásticos

Un *proceso estocástico* es una colección de variables aleatorias  $\{X_t, t \in T\}$  definidas todas en el mismo espacio muestral. Para nuestro propósito  $t$  tendrá connotaciones de tiempo. El conjunto de todos los posibles valores que puede tomar  $X_t$ , para cualquier valor de  $t$ , se llama espacio de estados del proceso estocástico. Si  $T$ , conjunto de todos los valores que puede tomar el índice  $t$ , es contable (normalmente los enteros positivos), entonces  $\{X_t, t \in T\}$  se llama *proceso estocástico de tiempo discreto*. Si  $T$  es un subconjunto incontable de los números reales (normalmente los reales no negativos),  $\{X_t, t \in T\}$  se llama *proceso estocástico de tiempo continuo*.

En el ejemplo de la cola atendida por un empleado, con intervalos IID entre llegadas,  $A_1, A_2, \dots$  y tiempos de servicio,  $S_1, S_2, \dots$ , puede definirse el proceso estocástico de tiempo discreto "tiempo de espera en la cola",  $\{D_i, i \geq 1\}$ , de la forma:

$$D_1 = 0$$

$$D_{i+1} = \max\{D_i + S_i - A_{i+1}, 0\} \quad \text{para } i = 1, 2, \dots$$

donde, como vemos, el tiempo de espera del cliente  $i+1$  en la cola es igual al tiempo que ha compartido con el cliente  $i$  en la cola más el tiempo de atención,  $S_i$ , al cliente  $i$ . El tiempo que el cliente  $i$  y el  $i+1$  comparten en la cola es igual al tiempo de espera del cliente  $i$ ,  $D_i$ , menos el tiempo que transcurrió entre que llegó el cliente  $i$  y el  $i+1$ ,  $A_{i+1}$ . En este caso el conjunto,  $T$ , de valores que puede tomar el índice  $t$  son los enteros positivos y el espacio de estados son los reales no negativos.

En ocasiones, para poder inferir acerca del proceso estocástico que da lugar a los datos de salida de la simulación, debemos hacer suposiciones acerca del proceso estocástico que pueden no ser estrictamente ciertas en la práctica y, sin las cuales, el análisis estadístico de los datos de salida no sería posible.

Un ejemplo de esto es suponer que un *proceso estocástico es de covarianza estacionaria*. Un proceso estocástico discreto  $\{X_i, i \geq 1\}$  se dice de covarianza estacionaria si:

$$\begin{cases} \mu_i = \mu & \text{para } i = 1, 2, \dots \quad \text{y } -\infty < \mu < \infty \\ \sigma_i^2 = \sigma^2 & \text{para } i = 1, 2, \dots \quad \text{y } \sigma^2 < \infty \\ C_{i,j+1} = \text{Cov}(X_i, X_{i+j}) & \text{independiente de } i, \text{ para } j = 1, 2, \dots \end{cases}$$

es decir, para un proceso estocástico de covarianza estacionaria, la media y la varianza son estacionarias en el tiempo (la media y la varianza comunes en el tiempo se han notado  $\mu, \sigma^2$ ) y la covarianza entre dos observaciones  $X_i$  y  $X_{i+j}$  depende sólo de la separación  $j$  y no de los valores de  $i$  e  $i+j$ . También se definen los procesos estocásticos continuos de covarianza estacionaria.

Para un proceso de covarianza estacionaria, notamos la covarianza entre  $X_i$  y  $X_{i+j}$  como  $C_j$  y la correlación como  $\rho_j$ , donde:

$$\rho_j = \frac{C_{i,i+j}}{\sqrt{\sigma_i^2 \sigma_{i+j}^2}} = \frac{C_j}{\sigma^2}$$