Universidad de Chile
Facultad De Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática
MA3403 - Probabilidades y Estadística

Vectores Aleatorios

Alberto Vera Azócar, albvera@ing.uchile.cl

1. Vectores Aleatorios bi-variados

Nuestro interés está en variables aleatorias X e Y definidas en un mismo espacio Ω, trabajaremos con vectores bi-variados, las generalizaciones a n variables no son difíciles de extender y por tanto quedan propuestas.

Definición 1.1. *Diremos que el par (X,Y) es un vector aleatorio si X e Y representan variables aleatorias*

La función que caracteriza el comportamiento del vector aleatorio es la distribución conjunta, ya que entrega la información de como se comporta el vector en cada componente y además el comportamiento cuando interactúan las componentes (uno pensaría que es más difícil obtener $\mathbb{P}(X > 2 \wedge Y < 1)$ que $\mathbb{P}(X > 2)$).

Definición 1.2. *Sean $X, Y : \Omega \to \mathbb{R}$ variables aleatorias, se define la distribución conjunta como*

$$F_{XY}(x, y) := \mathbb{P}(X \leq x, Y \leq y)$$

Las leyes marginales se refieren siempre a como se comporta una v.a. sin tener en cuenta el resultado de las demás, así tenemos:

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \mathbb{P}(X \leq x, Y < \infty) \\
&= \lim_{n\to\infty} \mathbb{P}(X \leq x, Y \leq n) \\
&= \lim_{n\to\infty} F_{XY}(x, n)
\end{aligned}$$

Definición 1.3. *Sean $X, Y : \Omega \to \mathbb{R}$ v.a., definimos las distribuciones marginales como*

$$F_X(x) := \lim_{y\to\infty} F_{XY}(x, y) \qquad F_Y(y) := \lim_{x\to\infty} F_{XY}(x, y)$$

1.1. Caso bi-discreto

Definición 1.4. *Diremos que (X,Y) es un vector discreto si $X, Y : \Omega \to \mathbb{R}$ son v.a. discretas, es decir si existen $A_x, A_y \subsetneq \mathbb{R}$ a lo más numerables tales que $\mathbb{P}(X \in A_x) = 1$ y $\mathbb{P}(Y \in A_y) = 1$.*

De la definición es claro que también existe un conjunto A a lo más numerable tal que $\mathbb{P}((X, Y) \in A) = 1$, pues basta tomar por ejemplo $A = A_x \times A_y$.

Se busca caracterizar al vector (X, Y), en general $F_{XY}(x, y)$ es suficiente, pero no es útil para los cálculos, no así como la ley conjunta:

Definición 1.5. *Sea (X, Y) un vector aleatorio discreto, se define la ley conjunta como*

$$p_{ij} := \mathbb{P}(X = i, Y = j)$$

1.1.1. Leyes Marginales

Las leyes marginales que denotaremos $p_i := \mathbb{P}(X = i)$ se obtienen particionando el espacio: dado que los eventos $Y = j$ y $Y = j+1$ son disjuntos $\forall j$ (una v.a. no puede tomar dos valores a la vez) y recordando que dada una partición $\{A_i\}_{i\in\mathbb{N}}$ del conjunto A se tiene

$$\mathbb{P}(A) = \mathbb{P}\left(\dot{\bigcup_{i\in\mathbb{N}}} A_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$$

entonces las leyes marginales se escriben como

$$p_i = \sum_{j\in\mathbb{N}} \mathbb{P}(X = i, Y = j) = \sum_{j\in\mathbb{N}} p_{ij} \tag{1}$$

$$p_j = \sum_{i\in\mathbb{N}} \mathbb{P}(X = i, Y = j) = \sum_{i\in\mathbb{N}} p_{ij}$$

1.1.2. Distribuciones Marginales

Dadas las leyes marginales, se obtienen directamente las distribuciones marginales:

$$\begin{aligned} F_X(x) &\triangleq \mathbb{P}(X \leq x) \\ &= \sum_{i\leq x} \mathbb{P}(X = i) \\ &= \sum_{i\leq x}\sum_{j\in\mathbb{N}} p_{ij} \qquad (2) \end{aligned}$$

donde en (2) se utilizó (1). Análogamente se tiene

$$F_Y(y) = \sum_{j\leq y}\sum_{i\in\mathbb{N}} p_{ij}$$

Siguiendo con el razonamiento de particionar, la función de distribución conjunta se calcula como

$$F_{XY}(x,y) = \sum_{i\leq x}\sum_{j\leq y} p_{ij}$$

1.2. Caso bi-continuo

Definición 1.6. *Diremos que $(X,Y) : \Omega \to \mathbb{R}^2$ es un vector conjuntamente continuo si existe una función $f_{XY} : \mathbb{R}^2 \to \mathbb{R}_+$ que llamaremos densidad conjunta tal que $\forall A \subseteq \mathbb{R}^2$ se tiene*[1]

$$\mathbb{P}((X,Y)\in A) = \int\int_A f_{XY}(x,y)dxdy$$

Las deducciones en el caso discreto las realizamos con argumentos de partición, usando implícitamente la σ-aditividad, que es válida para uniones numerables. Como en el caso continuo no podemos hacer esto las deducciones se harán vía derivación.

[1] La definición formalmente no es para todo conjunto, sino para aquellos que pertenezcan a una familia bastante general. Es válida sí para cualquier conjunto A que nos interese en el curso.

Propiedades

1. Claramente la integral en todo el soporte da 1, el caso general es

$$\int\int_{\mathbb{R}^2} f_{XY}(x,y)dxdy = 1$$

2. De la definición, eligiendo $A = (-\infty, x) \times (-\infty, y)$

$$F_{XY}(x,y) \triangleq \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)dxdy$$

3. Derivando F_{XY} se puede obtener la densidad conjunta

$$\frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) = f_{XY}(x,y)$$

1.2.1. Densidades Marginales

Sabemos que

$$\begin{aligned} F_X(x) &\triangleq \lim_{y\to\infty} F_{XY}(x,y) \\ &= \lim_{y\to\infty} \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy \end{aligned}$$

Usando el Teorema Fundamental del Cálculo

$$\frac{d}{dx}F_X(x) = \frac{d}{dx}\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

Por lo anterior parece natural definir

$$f_X(x) := \int_{\mathbb{R}} f_{XY}(x,y)dy \qquad f_Y(y) := \int_{\mathbb{R}} f_{XY}(x,y)dx$$

2. Independencia

Definición 2.1. *Diremos que X e Y v.a. son independientes si cumplen que para cualquier par de conjuntos A y B se tiene*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

La definición es difícil de trabajar, por lo que se hace necesario tener más caracterizaciones de independencia. Aceptaremos el siguiente Teorema sin demostración.

Teorema 2.1. *Sea (X,Y) un vector aleatorio, las v.a. X e Y son independientes ssi*
1) $F_{XY}(x,y) = F_X(x)F_Y(y)$
2) Es vector bi-discreto y $p_{ij} = p_i p_j$
3) Es vector bi-continuo y $f_{XY}(x,y) = f_X(x)f_Y(y)$

Las caracterizaciones 2 y 3 son las más útiles, pero la 1 es válida para casos generales, por ejemplo X continua e Y discreta.

3. Esperanza de Una Función

Consideremos una función $g : \mathbb{R}^2 \to \mathbb{R}^m$, nos interesa calcular su valor esperado respecto a la ley (X, Y), en particular cuando escojamos a g como la identidad estaremos calculando la esperanza del vector.

Veremos las definiciones para un vector bi-discreto, los otros casos son fáciles de extender y quedan propuestos.

Definición 3.1. *Sea (X,Y) un vector aleatorio discreto y $g : \mathbb{N}^2 \to \mathbb{R}$ función a valor real. Definimos la esperanza de g como*

$$\mathbb{E}(g(X,Y)) := \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} g(i,j) p_{ij} \tag{3}$$

Observamos que la esperanza en el caso anterior es un escalar, para el caso en que el co-dominio de g tenga mayor dimensionalidad que 1 se tiene:

Definición 3.2. *Sea (X,Y) un vector aleatorio discreto y $g : \mathbb{N}^2 \to \mathbb{R}^m$, se define*

$$\mathbb{E}(g(X,Y)) := (\mathbb{E}(g_1(X,Y)), \mathbb{E}(g_2(X,Y)), \cdots, \mathbb{E}(g_m(X,Y)))$$

La esperanza ("a secas") del vector (X,Y) es mediante la función $g(x,y) = (x,y)$, con lo que

$$\mathbb{E}((X,Y)) := (\mathbb{E}(X), \mathbb{E}(Y))$$

La esperanza inducida en $\mathbb{E}(X)$ puede ser por ley marginal p_i o bien por la conjunta p_{ij}, en efecto, utilizando la definición (3) antes vista:

$$\mathbb{E}(X) \triangleq \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} i p_{ij} = \sum_{i\in\mathbb{N}} i \sum_{j\in\mathbb{N}} p_{ij} = \sum_{i\in\mathbb{N}} i p_i$$

4. Función Generadora de Momentos

La función generadora de momentos nos será útil como herramienta para identificar distribuciones, se encuentran tabuladas las funciones generadoras de muchas funciones conocidas (Poisson, exponencial, etc), entonces para encontrar la distribución de una nueva variable aleatoria a veces puede ser conveniente calcular su función generadora para concluir utilizando las que son conocidas.

Definición 4.1. *Sea X una v.a. y $n \in \mathbb{N}$, se define el momento de orden n de X como $\mathbb{E}(X^n)$.*

Definición 4.2. *Sea X una v.a., se define la función generadora de momentos de X como*

$$\psi_X(t) := \mathbb{E}(e^{tX}) \tag{4}$$

el dominio de ψ_X es t tal que la esperanza en (4) converge.

Propiedades:

1. Sea X v.a., si ψ_X es k veces derivable, entonces

$$\left.\frac{d^k}{dt^k}\psi_X(t)\right|_{t=0} = \mathbb{E}(X^k)$$

2. Sean X e Y v.a., se tiene que

$$\psi_Y(t) = \psi_X(t) \Rightarrow f_X(x) = f_Y(y)$$

3. Sean X e Y v.a., se tiene que

$$X \perp\!\!\!\perp Y \Rightarrow \psi_{X+Y}(t) = \psi_X(t) \cdot \psi_Y(t)$$

5. Teorema del Cambio de Variables

Teorema 5.1. *Sea (X,Y) un par conjuntamente continuo con densidad f_{XY}, sea $\mathcal{U} \subseteq \mathbb{R}^2$ abierto tal que $\mathbb{P}((X,Y) \in \mathcal{U}) = 1$. Sea $\varphi : \mathcal{U} \to \varphi(\mathcal{U})$ función $\in \mathcal{C}^1(\mathcal{U})$ biyectiva con $det(J(\varphi)) \neq 0$. Entonces si denotamos $(Z,W) = \varphi(X,Y)$*

$$f_{ZW}(z,w) = f_{XY}(\varphi^{-1}(z,w))|det(J(\varphi^{-1}))|\mathbb{1}_{\{(z,w)\in\varphi(\mathcal{U})\}}$$

6. Covarianza y Correlación

No siempre es fácil estudiar la independencia, por eso usaremos conceptos que tomarán mayor relevancia en la parte estadística, que además en sus versiones empíricas son los verdaderos objetos de estudio al momento de ver independencia.

Definición 6.1. *Sean $X, Y : \Omega \to \mathbb{R}$ v.a. que cumplen $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, se define la covarianza entre X e Y como*

$$cov(X,Y) := \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

Bajo los supuestos de la definición la covarianza está bien definida ($cov(X,Y) < \infty$), ahora se puede calcular como esperanza de una función, por ejemplo si (X,Y) es un vector conjuntamente continuo, entonces

$$cov(X,Y) = \int\int_{\mathbb{R}^2} (x - \mathbb{E}(X))(y - \mathbb{E}(Y))f_{XY}(x,y)dxdy$$

La covarianza es una medida de relación entre las variables, es adecuado pensar que si la covarianza es muy grande entonces las v.a. están fuertemente relacionadas; si es muy negativa por ejemplo se intuye que las v.a. están muy relacionadas y que además cuando una aumenta el valor la otra disminuye.

Para medir distintas variables, digamos X, Y, Z la covarianza puede ser engañosa por cuanto depende de la unidad de medida de las variables, por esto se "estandariza" como el coeficiente de correlación (lineal), este sí es comparable entre distintos pares de variables.

Definición 6.2. *Sean $X, Y : \Omega \to \mathbb{R}$ v.a. con $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, se define el coeficiente de correlación como*

$$\rho(X,Y) := \frac{cov(X,Y)}{\sqrt{\mathbb{V}ar(X)}\sqrt{\mathbb{V}ar(Y)}}$$

El coeficiente de correlación cumple $-1 \leq \rho(X,Y) \leq 1$.

La propiedad más famosa de la covarianza es

$$X \perp\!\!\!\perp Y \Rightarrow cov(X,Y) = 0$$

Es importante recalcar que la implicancia es sólo hacia la derecha, $cov(X,Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$.

7. Condicionalidad

Ya se ha visto que la probabilidad condicional induce una medida de probabilidad, es decir, dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ si tomamos a $B \subseteq \Omega$ con $\mathbb{P}(B) > 0$, entonces $\mathbb{P}(\cdot|B)$ es medida de probabilidad, donde:

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Lo que ahora nos motiva es reemplazar el evento B por el evento $Y = y$ donde Y es una v.a., en el caso discreto esto es más natural que en el caso continuo, pero luego uno se convence.

7.1. Caso Discreto

Definición 7.1. *Sean $X, Y : \Omega \to \mathbb{R}$ v.a. discretas, se define la probabilidad de que $X = i$ dado que $Y = j$ cuando $\mathbb{P}(Y = j) > 0$ como*

$$\mathbb{P}(X = i|Y = j) := \frac{\mathbb{P}(\{X = i, Y = j\})}{\mathbb{P}(Y = j)}$$

Dada la definición anterior consideraremos como ley condicional a la colección de todos los valores de probabilidad condicional, denotaremos para $\mathbb{P}(Y = y) > 0$ a la ley condicional de X dado Y como

$$p_{X|Y}(x|y) := \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

Observamos que para calcular la ley condicional necesitamos la ley conjunta p_{ij}, en efecto es fácil ver que

$$p_{X|Y}(i|j) = \frac{p_{ij}}{\sum_{k\in\mathbb{N}} p_{kj}}$$

Si no tenemos la ley conjunta $\mathbb{P}(X = i, Y = j)$, pero tenemos la ley marginal de Y y la ley condicional de X dado Y, entonces podemos obtener la ley conjunta y por lo tanto la marginal de X, en efecto

$$\begin{aligned} \mathbb{P}(X = i, Y = j) &= \mathbb{P}(X = i|Y = j)\mathbb{P}(Y = j) \\ \Rightarrow \mathbb{P}(X = i) &= \sum_{j\in\mathbb{N}} \mathbb{P}(X = i|Y = j)\mathbb{P}(Y = j) \end{aligned}$$

Definiremos también a la distribución condicional como

$$F_{X|Y}(x|y) := \mathbb{P}(X \leq x|Y = y) = \sum_{i\leq x} p_{X|Y}(i|y)$$

7.2. Caso Continuo

En este caso es un poco raro pensar en el observable $Y = y$, ya que si Y es v.a. continua, entonces $\forall y \in \mathbb{R}, \quad \mathbb{P}(Y = y) = 0$, la intuición es que si bien eso es cierto cuando uno echa a correr una v.a. igual se obtiene un resultado (alguna temperatura se observa por ejemplo) y este hecho da información sobre la ocurrencia de la v.a. X.

Definición 7.2. *Sean $X, Y : \Omega \to \mathbb{R}$ v.a. conjuntamente continuas, es decir, que tienen densidad conjunta f_{XY}, se define la densidad condicional de X dado el observable $Y = y$ como*

$$f_{X|Y}(x|y) := \frac{f_{XY}(x, y)}{f_Y(y)} \mathbb{1}_{\{f_Y(y)>0\}}$$

Además definiremos para A un conjunto unión de intervalos la probabilidad de que $X \in A$ dado un observable como

$$\mathbb{P}(X \in A|Y = y) := \int_A f_{X|Y}(x|y)dx$$

La principal herramienta que tendremos para cálculo mediante probabilidades condicionales es el siguiente Teorema (la demostración es un poco delicada, pero no es muy difícil):

Teorema 7.1. *Sean $X, Y : \Omega \to \mathbb{R}$ v.a. conjuntamente continuas, para A conjunto se tiene que*

$$\mathbb{P}(X \in A) = \int_{\mathbb{R}} \mathbb{P}(X \in A|Y = y)f_Y(y)dy$$

7.3. Esperanza Condicional

Tendremos la esperanza definida análogamente que para el caso conjunto, solo que se reemplaza ley (densidad) conjunta por la condicional, así dada una función $g : \mathbb{R} \to \mathbb{R}$ se tiene en el caso discreto

$$\mathbb{E}(g(X)|y) := \sum_{i \in \mathbb{N}} g(i) p_{X|Y}(i|y)$$

y en el caso continuo

$$\mathbb{E}(g(X)|Y) := \int_{\mathbb{R}} g(x) f_{X|Y}(x|y) dx$$

Ya hemos visto varias herramientas, en particular se vio que dada la ley condicional se puede obtener la marginal, así para calcular $\mathbb{E}(X)$ se puede primero obtener la marginal y luego aplicar definición de esperanza, a veces esto complica, por lo que se tiene el siguiente Teorema que nos ayuda a calcular la esperanza de X dada la condicional (válido para v.a. continuas y discretas).

Teorema 7.2. *Sean $X, Y : \Omega \to \mathbb{R}$ v.a., se tiene que*

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$$