2 Repaso de Variables Aleatorias y Funciones de Distribución

2.1 Conceptos familiares.

El objetivo es presentar un breve repaso de las nociones de probabilidad, necesarias para iniciar el estudio de procesos estocásticos.

Una variable aleatoria (v.a.) X es llamada ***discreta*** si hay un conjunto de valores distintos, finito o numerable x_1, x_2, ..., tales que $p_i = P\{X = x_i\} > 0$, i = 1, 2, 3, ..., y $\sum_i p_i = 1$. Si P{X = x} = 0 para cada valor de x, la variable aleatoria es llamada ***continua***. Si hay una función no negativa f(t), definida para $-\infty < t < \infty$ tal que la ***función de distribución F*** sobre la variable aleatoria X es dada por

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

entonces se dice que ***f es la densidad de probabilidad*** de X. Si X tiene una densidad de probabilidad, entonces necesariamente es continua, pero hay ejemplos de variables aleatorias las cuales no poseen densidades de probabilidad (**Tarea**. Buscar ejemplos de variables aleatorias que no posean densidad de probabilidad).

La cantidad $E[X^n]$, $n \geq 1$, es llamada el ***n-ésimo momento*** de X. Se define como:

$$E\left[X^n\right] = \begin{cases} \sum_{x:p(x)>0} x^n \; p_X(x) & si\;\; X\; es\, discreta \\ \int_{-\infty}^{\infty} x^n \; f_X(x)\; dx & si\;\; X\; es\, continua. \end{cases}$$

El primer momento de X es conocido como la ***media***, y es usual denotarla por μ_X. El ***m-ésimo momento central*** de X se define como el m-ésimo momento de la variable aleatoria $X-\mu_X$ si μ_X existe. El primer momento central es evidentemente cero; el segundo momento central es llamado la ***varianza*** σ_X^2 de X. La ***mediana*** de una variable aleatoria X es cualquier m con la propiedad de que $P\{X \geq m\} \geq ½$ y $P\{X \leq m\} \geq ½$.

Si X es una variable aleatoria y g es una función, entonces Y = g(X) es también una variable aleatoria. Su valor esperado está dado por

$$E[g(X)] = \begin{cases} \sum_{i=1}^{\infty} g(x_i) P\{X = x_i\}, & si\ g(X)\ es\, discreta \\ \int g(x)\, f_X(x)\, dx, & si\ g(X)\, es\, continua. \end{cases}$$

2.2 Funciones de distribución conjunta.

Dado un par (X,Y) de variables aleatorias, su ***función de distribución conjunta*** es la función F_{XY} de dos variables reales dadas por

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

La función $F(x,+\infty) \equiv \lim_{y\to\infty} F(x,y)$ es una función de distribución de probabilidad llamada la ***función de distribución marginal de X***. De manera similar, la función $F(+\infty,y)$ es llamada la ***distribución marginal de Y***. Si sucede que $F(x,+\infty)\cdot F(+\infty,y) = F(x,y)$ para cada elección de x, y, entonces las variables aleatorias X y Y son independientes. Una función de distribución conjunta F_{XY} se dice que posee una densidad de probabilidad conjunta si existe una función $f_{XY}(s,t)$ de dos variables reales tal que

$$F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(s,t)\; ds\, dt$$

para todo x, y. Si X y Y son independientes, entonces $f_{XY}(s,t)$ es necesariamente de la forma $f_X(s)\cdot f_Y(t)$, donde f_X y f_Y son las densidades de probabilidad de la distribución marginal de X y Y, respectivamente.

La función de distribución conjunta de cualquier colección finita $X_1, ..., X_n$ de variables aleatorias es definida como la función

$$\begin{aligned} F(x_1, ..., x_n) &= F_{X1,...,Xn}(x_1, ..., x_n) \\ &= P[X_1 \leq x_1, ..., X_n \leq x_n\} \end{aligned}$$

La función de distribución

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \lim_{x_i \to \infty,\, i \neq i_1,\ldots,i_k} F(x_1, \ldots, x_n)$$

es llamada la distribución marginal de las variables aleatorias $X_{i_1},\ldots,X_{i_k}$. Si $F(x_1, \ldots, x_n) = F_{X_1}(x_1)\cdot\ldots\cdot F_{X_n}(x_n)$ para todos los valores de $x_1, \ldots, x_n$, las variables aleatorias $X_1, \ldots, X_n$ se dice que son independientes.

Una función de distribución conjunta $F(x_1, \ldots, x_n)$ se dice que tiene una densidad de probabilidad si existe una función no negativa $f(t_1, \ldots, t_n)$ de n variables tal que

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1} f(t_1,\ldots,t_n)\, dt_1 \cdots dt_n$$

para todo real $x_1, \ldots, x_n$.

Si X y Y son variables aleatorias distribuidas conjuntamente teniendo medias μ_X y μ_Y, respectivamente, su ***covarianza*** σ_{XY} es el momento producto

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)].$$

Si X_1 y X_2 son variables aleatorias independientes teniendo las funciones de distribución F_1 y F_2, respectivamente, entonces la función de distribución F de la suma $X = X_1 + X_2$ es la ***convolución*** de F_1 y F_2:

$$F(x) = \int F_1(x-y)\, dF_2(y) = \int F_2(x-y)\, dF_1(y).$$

Especializándose a la situación donde X_1 y X_2 tienen las densidades de probabilidad f_1 y f_2, la función de densidad f de la suma $X = X_1 + X_2$ es la ***convolución*** de las densidades f_1 y f_2:

$$f(x) = \int f_1(x-y)\, df_2(y) = \int f_2(x-y)\, df_1(y).$$

2.3 Distribuciones condicionales y esperanzas condicionales.

La ***probabilidad condicional*** $P\{A|B\}$ del evento A dado el evento B es definido por

$$P\{A \mid B\} = \frac{P\{A\, y\, B\}}{P\{B\}}, \quad \text{si } P\{B\} > 0,$$

y es indefinida a la izquierda, o asignado un valor arbitrario, cuando P{B} = 0.

Sean X y Y variables aleatorias las cuales pueden obtener sólo una infinidad de valores numerables, digamos 1, 2, ... La ***función de distribución condicional*** $F_{X/Y}(\cdot / y)$ de X dado Y = y se define por

$$F_{X/Y}(\cdot \mid y) = \frac{P\{X \le x, Y = y\}}{P\{Y = y\}}, \quad \text{si } P\{Y = y\} > 0,$$

Supongamos ahora que X y Y son variables aleatorias continuas distribuidas conjuntamente teniendo la función de densidad de probabilidad conjunta $f_{XY}(x,y)$. Entonces la ***distribución condicional*** de X dado Y = y es dada por

$$F_{X|Y}(x \mid y) = \frac{\int_{t \le x} f_{XY}(t, y)\, dt}{f_Y(y)}, \quad \text{siempre que } f_Y(y) > 0.$$

$F_{X/Y}(x / y)$ satisface:

i. $F_{X/Y}(x / y)$ es una función de distribución de probabilidad en x para cada y fijo.

ii. $F_{X/Y}(x / y)$ es una función de y para cada x fija; y

iii. Para cualesquiera valores x, y

$$P\{X \le x, Y \le y\} = \int_{\eta \le y} F_{X|Y}(x \mid \eta)\, dF_Y(\eta)$$

Donde $F_Y(\eta) = P\{Y \le \eta\}$ es la distribución marginal de Y.

Cuando Y es una variable aleatoria continua que tiene la función de densidad de probabilidad $f_Y(y)$, la integral definida en el último punto es calculada como

$$P\{X \le x, Y \le y\} = \int_{\eta \le y} F_{X|Y}(x \mid \eta)\, f_Y(y)\, d\eta .$$

Y cuando Y es discreta, la fórmula es

$$P\{X \le x, Y \le y\} = \sum_{\eta \le y} F_{X|Y}(x \mid \eta)\, P\{Y = \eta\} .$$

Esas tres propiedades (i-iii) capturan la característica esencial de las distribuciones condicionales. De hecho, a partir del tercer punto obtenemos

$P\{X \le x, Y = y\} = P\{X \le x, Y \le y\} - P\{X \le x, Y < y\}$

$$= \sum_{\eta \le y} F_{X|Y}(x \mid \eta)\ P\{Y = \eta\} - \sum_{\eta < y} F_{X|Y}(x \mid \eta)\ P\{Y = \eta\}$$

$$= F_{X|Y}(x \mid y)\, P\{Y = y\}$$

lo cual entonces implica la definición $F_{X|Y}(x \mid y) = P\{X \le x, Y \le y\}\, P\{Y = y\}$, en donde P{Y=y} > 0. de igual modo, esas tres propiedades son tomadas como la base para la definición de las distribuciones condicionales.

La aplicación de (iii) en el caso y = ∞ produce la ***Ley de Probabilidad Total***

$$P\{X \le x\} = P\{X \le x, Y < \infty\} = \int_{-\infty}^{+\infty} F_{X|Y}(x \mid y)\, dF_Y(y),$$

la cual es una de las fórmulas más fundamentales del análisis de probabilidad.

Cuando Y es discreta esta relación es

$$P\{X \le x\} = \sum_{y} P\{X \le x \mid Y \le y\}\, P\{Y \le y\}.$$

Cuando Y tiene función de densidad de probabilidad $f_Y(y)$ se tiene

$$P\{X \le x\} = \int_{-\infty}^{+\infty} P\{X \le x | Y = y\}\ f_Y(y)\, dy.$$

Cuando X y Y son variables aleatorias continuas distribuidas conjuntamente, podemos definir la ***función de densidad condicional*** $f_{X|Y}(x \mid y)$ de X dado que Y = y por

$$f_{X|Y}(x \mid y) = \frac{d}{dx} F_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

en valores de y para los cuales $f_Y(y) > 0$.

Sea g una función para la cual la esperanza de g(X) es finita. La ***esperanza condicional*** de g(X) dado que Y = y puede ser expresada en la forma

$$E[g(X)\,|\,Y=y]=\int_x g(x)\,dF_{X|Y}(x\,|\,y)\,.$$

Cuando X y Y son variables aleatorias continuas conjuntas, E[g(X)|Y = y] puede ser calculada de

$$E[g(X)\,|\,Y=y]=\int g(x)\,f_{X|Y}(x\,|\,y)\,dx$$

$$=\frac{\int g(x)\,f_{XY}(x,y)\,dx}{f_Y(y)}\,, \qquad \text{si } f_Y(y)>0,$$

y si X y Y son variables aleatorias discretas distribuidas conjuntamente, tomando los posibles valores x_1, x_2, ..., entonces la fórmula se reduce a

$$E[g(X)\,|\,Y=y]=\sum_{i=1}^{\infty} g(x_i)\,P[X=x_i\,|\,Y=y]$$

$$=\frac{\sum_{i=1}^{\infty} g(x_i)\,P\{X=x_i, Y=y\}}{P\{Y=y\}}\,, \qquad \text{si } P\{Y=y\}>0.$$

En paralelo a las propiedades (i-iii) vemos que la esperanza condicional de g(X) dado Y = y satisface

(E.1) E[g(X)|Y = y] es una función de y para cada función g para la cual E[|g(X)|] < ∞; y

(E.2) Para cada función acotada h tenemos

$$E[g(X)h(Y)] = \int E[g(X)|Y = y]\, h(y)\, dF_Y(y)$$

donde F_Y es la función de distribución marginal para Y.

El caso especial en (E.2) con h(y) ≡ 1 produce la fórmula que expresa la Ley de Probabilidad Total para esperanzas.

$$E[g(X)h(Y)] = \int E[g(X)|Y = y]\; dF_Y(y),$$

la cual, cuando Y es discreta, se tiene

$$E[g(X)] = \sum_{i=1}^{\infty} E[g(X)\,|\,Y=y_i]\,P\{Y=y_i\}\,,$$

y, cuando Y tiene función de densidad f_Y, será

$$E[g(X)] = \int E[g(X) \mid Y = y]\; f_Y(y)\, dy .$$

Puesto que la esperanza condicional de g(X) dado que Y = y es la esperanza con respecto a la distribución condicional $F_{X|Y}$, la esperanza condicional se comporta en cierta forma igual que las esperanzas ordinarias. En particular, si a_1 y a_2 son números fijos y g_1 y g_2 son funciones dadas para las cuales $E[|g_i(X)|] < \infty$, i = 1, 2, entonces

$$E[a_1 g_1(X) + a_2 g_2(X) \mid Y = y] = a_1 E[g_1(X) \mid Y = y] + a_2 E[g_2(X) \mid Y = y]$$

De acuerdo a (E.1), E[g(X)|Y = y] es una función de la variable real y. Si evaluamos esta función en la variable aleatoria Y, obtenemos una variable aleatoria la cual denotamos por E[g(X)|Y]. La propiedad básica (E.2) es entonces establecida para cualquier función acotada h de y.

E[g(X) h(Y)] = E{E[g(X)|Y] h(Y)}

Cuando h(y) = 1 para todo y, se obtiene la ley de probabilidad total en la forma

E[g(X)] = E{E[g(X)|Y}.

La siguiente lista resume todas las propiedades de esperanza condicional. Aquí con o sin subíndices, X y Y son variables aleatorias, c es un número real, g es una función para la cual $E[|g(X)|] < \infty$, f es una función acotada y h es una función de dos variables para la cual $E[|h(X, Y)|] < \infty$.

a) $E[a_1 g_1(X) + a_2 g_2(X) \mid Y] = a_1 E[g_1(X) \mid Y] + a_2 E[g_2(X) \mid Y]$,
b) $g \geq 0$ implica que $E[g(X)|Y] \geq 0$,
c) E[h(X, Y)|Y = y] = E[h(X, y)|Y =y],
d) E[g(X)|Y] = E[g(X)] si X y Y son independientes,
e) E[g(X)f(Y)|Y] = f(Y) E[g(X)|Y],
f) E[g(X)f(Y)] = E{ f(Y) E[g(X)|Y]},
g) E[c |Y] = c , cuando $g \equiv 1$,
h) E[f(Y)| Y] = f(Y),
i) E[g(X)] = E{E[g(X)|Y]}, cuando $f \equiv 1$.

2.4 Teoremas Límites

Sean Z, Z_1, Z_2, ... variables aleatorias distribuidas conjuntamente.

A. *Convergencia con probabilidad uno.*

Decimos que Z_n converge a Z con probabilidad uno si $P\{\lim_{n\to\infty} Z_n = Z\} = 1$.

En palabras, $\lim_{n\to\infty} z_n = z$, para un conjunto de resultados Z = z, $Z_1 = z_1$, $Z_2 = z_2$, ... teniendo probabilidad total igual a uno.

B. *Convergencia en probabilidad.*

Decimos que Z_n converge a Z en probabilidad si para cada ε positivo

$\lim_{n\to\infty} P\{|Z_n - Z| > \varepsilon\} = 0$,

o recíprocamente, si para cada ε positivo

$\lim_{n\to\infty} P\{|Z_n - Z| \leq \varepsilon\} = 1$.

En palabras, tomando n suficientemente grande, uno puede alcanzar arbitrariamente probabilidades altas de que Z_n sea arbitrariamente cercano a Z.

C. *Convergencia en media cuadrática.*

Decimos que Z_n converge a Z en media cuadrática si

$\lim_{n\to\infty} E\left[|Z_n - Z|^2\right] = 0$.

D. *Convergencia en distribución.*

Sea F(t) = P{Z ≤ t} y $F_k(t) = P\{Z_k \leq t\}$, k = 1, 2, ... Decimos que Z_n converge en distribución a Z si el $\lim_{n\to\infty} F_n(t) = F(t)$

Para toda t para la cual F es continua.

Nota. Puede probarse que si Z_n converge a Z con probabilidad uno, entonces Z_n converge a Z en probabilidad, y que esto implica que Z_n converge a Z en distribución. Así, la convergencia en distribución es la forma más débil de convergencia.

Muchos de los resultados básicos de la teoría de probabilidad están en la forma de teoremas limites y se mencionan unos cuantos aquí.

Sean X_1, X_2, ... variables aleatorias idénticamente distribuidas e independientes con media finita m. Sea $S_n = X_1 + \cdots + X_n$ y sea $\overline{X}_n = S_n/n$ que es la media simple.

Ley de los Grandes Números (Débil) . $\overline{X}_n$ converge en probabilidad a m. Es decir, para cada $\varepsilon > 0$

$$\text{Lim } P\{|\overline{X}_n - m| > \varepsilon\} = 0.$$

- **Ley de los Grandes Números (Fuerte).** $\overline{X}_n$ converge a m con probabilidad uno. Es decir,

  $$P\{\lim_{n\to\infty} \overline{X}_n = m\} = 1$$

- **Teorema del Limite Central**. Supongamos que cada X_k tiene varianza finita σ^2. Sea

  $$Z_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{1}{\sigma}(\overline{X}_n - m)\sqrt{n},$$

y sea Z una variable aleatoria normalmente distribuida teniendo media cero y varianza uno. Entonces Z_n converge en distribución a Z. Esto es, para todo real *a*,

$$\lim_{n\to\infty} P\{Z_n \le a\} = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du.$$

- **Lema de Borel- Cantelli**. Sean A_1, A_2, ... una sucesión infinita de eventos independientes tal que $P(A_1) + P(A_2) + \cdots < \infty$ y sea $B_n = A_n \cup A_{n+1} \cup \cdots$. Entonces

  $P(B_1 \cap B_2 \cap \cdots) = 0.$