

Resumen básico de materias de probabilidades

Roberto Alfredo Lineros Rodríguez
rlineros@puc.cl

1 de febrero de 2001

Índice

1. Introducción	4
2. Definiciones de Probabilidad	4
2.1. Definición clásica	4
2.2. Probabilidad “a priori”	4
2.3. Definición axiomática de probabilidad	4
3. Métodos de conteo	5
3.1. Variaciones	5
3.2. Permutaciones	6
3.3. Combinaciones	6
4. Probabilidad condicional	6
4.1. Regla Multiplicativa	7
4.2. Definición de Probabilidad Total y Teorema de Bayes	7
5. Variables Aleatorias	8
6. Modelos probabilísticos	8
6.1. Modelos discretos	9
6.1.1. Distribución Bernoulli	9
6.1.2. Distribución Binomial	9
6.1.3. Distribución geométrica	9
6.1.4. Distribución binomial negativa	9
6.1.5. Distribución hipergeométrica	10
6.1.6. Distribución Poisson	10
6.2. Modelos Continuos	10
6.2.1. Distribución Uniforme	10
6.2.2. Distribución Normal	10
6.2.3. Distribución Gamma	11
7. Vectores Aleatorios	12
7.1. Distribuciones marginales	12
7.2. Distribuciones condicionales	12
8. Transformación de variables aleatorias y vectores aleatorios	12
9. Distribuciones de varias variables aleatorias	13
9.1. Distribución Multinomial	13
9.2. Distribución Normal Bivariada	14
10. Esperanza o valor esperado	14
10.1. Esperanza de una función	15
10.2. Esperanza de un vector aleatorio	15

<i>ÍNDICE</i>	3
11.Desviación media, varianza y desviación standard	16
11.1. Desviación media	16
11.2. Varianza	16
11.3. Desviación Standard	16
12.Desigualdad de Chebyshev	16
13.Covarianza y Correlación	17
13.1. Covarianza	17
13.2. Coeficiente o índice de correlación	17
14.Esperanza y Varianza condicional	18

1. Introducción

Este texto es un resumen de las materias vistas en el ramo de probabilidades, en enero del 2001. El texto solo sirve de referencia o formulario, además no se tratan los temas muy profundamente, cualquier error que se encuentre por favor avisarme.

2. Definiciones de Probabilidad

2.1. Definición clásica

Esta definición de probabilidad viene dada por la forma de como es obtenida, mas bien resulta de realizar el experimento. Dicho de otra forma,

$$P(c) = \lim_{n \rightarrow \infty} \frac{n_c(n)}{n} \quad (1)$$

donde $P(c)$, representa la probabilidad de que un evento c ocurra, n es el numero de intentos realizados y $n_c(n)$ son los aciertos obtenidos hasta el intento n . El problema de esta definición, es que resulta poco practica, ya que necesitan muchos intentos para definir la probabilidad del evento c .

2.2. Probabilidad “a priori”

Básicamente se trata de una definición, que se propone antes de realizar el experimento, y suponiendo que los resultados que se esperan son equiprobables, o sea, tienen la misma opción de salir. Matemáticamente la probabilidad de que resulte un evento c , viene dada por

$$P(c) = \frac{\text{card}(c)}{\text{card}(\Omega)} \quad (2)$$

donde $\text{card}(X)$ se refiere a una cardinalidad del evento X , o sea, es una idea de “medida” del evento X , en este caso $\text{card}(c)$ se refiere a la cantidad de casos en donde se cumple el evento c , por otro lado $\text{card}(\Omega)$, son la cantidad de casos totales, ya que Ω representa el universo de casos posibles en el experimento.

2.3. Definición axiomática de probabilidad

Sea Ω espacio muestral, \mathcal{A} es un σ -álgebra y P función definida como,

$$\begin{aligned} P : \mathcal{A} &\rightarrow \mathbf{R} \\ A &\mapsto P(A) \end{aligned}$$

satisface que:

1. $P(A) \geq 0 \forall A \in \mathcal{A}$

2. $P(\Omega) = 1$
3. Si A_1 y A_2 son mutuamente excluyentes, o sea, $A_1 \cap A_2 = \phi$ entonces

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

En base a estos axiomas que definen a la probabilidad, se extraen estas propiedades:

- $P(\phi) = 0$ ya que ϕ representa es el conjunto vacío.
- Si A_i y A_j son mutuamente excluyentes, o sea, $A_i \cap A_j = \phi \forall i \neq j$ entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- $P(A^c) = 1 - P(A)$ donde A^c es el complemento de A , o sea $A^c \cup A = \Omega$ y además $A^c \cap A = \phi$.
- $0 \leq P(A) \leq 1$
- Sean A y B eventos cualesquiera, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Si $A \subseteq B$ con A y B eventos cualesquiera, entonces

$$P(A) \leq P(B)$$

- Sean A y B eventos cualesquiera, entonces

$$P(A \cup B) \leq P(A) + P(B)$$

3. Métodos de conteo

Los métodos de conteo se utilizan básicamente para encontrar los casos favorables y además es útil para saber la cantidad de casos totales.

3.1. Variaciones

Existen 2 tipos de variaciones, estos son variaciones sin repetición y con repetición.

$$V(m, n) = \frac{m!}{(m - n)!} \quad (3)$$

Donde (3) corresponde a la variación sin repetición, o sea, son las maneras posibles y distintas de seleccionar n objetos de m , importando el orden y sin que estos se repitan. Ahora bien, existe otro caso,

$$V^R(m, n) = m^n \quad (4)$$

donde (4) es una variación con repetición, esta vez, son las distintas maneras de seleccionar n objetos de m , importando el orden y pudiendo esto repetirse.

3.2. Permutaciones

Al igual que las variaciones existen también 2 tipos de permutaciones, estas son:

$$P^m = V(m, m) = m! \quad (5)$$

en este caso la ecuación (5) corresponde a una permutación sin reemplazo, que corresponde a todas las formas distintas en que pueden ordenarse m elementos distintos.

$$P_{n_1, \dots, n_k}^m = \binom{m}{n_1 \dots n_k} = \frac{m!}{n_1! \dots n_k!} \quad (6)$$

muy parecido a la permutación sin repetición, pero en este caso existen elementos repetidos, tal que se debe cumplir que, $\sum n_i = m$.

3.3. Combinaciones

Se definen 2 clase de combinaciones, combinación sin repetición,

$$C_n^m = \binom{m}{n} = \frac{m!}{(m-n)!n!} \quad (7)$$

corresponde a todas las formas distintas de seleccionar n objetos de un total de m , sin importar el orden, además son m elementos distintos.

$$C_{m,n}^R = \binom{m+n-1}{n} \quad (8)$$

este caso es igual que antes, pero es una combinación de elementos con reemplazo.

4. Probabilidad condicional

Cuando dos eventos A y B , están relacionados de alguna forma, tal que la probabilidad de alguno depende del otro, se puede hablar de probabilidad condicional, entonces para estos casos se introduce una nueva notación

$$P(A/B) \text{ probabilidad de } A \text{ dado } B$$

esto nos da a entender que se quiere la *probabilidad del evento A sabiendo que el evento B ocurrió* por otra parte

$P(A)$ probabilidad de A

que como se dijo es la probabilidad de A , pero esta vez sin considerar al evento B .

Definición: Sean A y B eventos cualesquiera con $P(B) > 0$, se define

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (9)$$

como la probabilidad condicional de A dado B .

4.1. Regla Multiplicativa

Sean A y B eventos cualesquiera, y además independientes entre si, entonces

$$P(A \cap B) = P(A)P(B) \quad (10)$$

esto ultimo se conoce como *regla multiplicativa*.

4.2. Definición de Probabilidad Total y Teorema de Bayes

Sea Ω espacio muestral, y existen también eventos E_i , tal que, $E_i \subseteq \Omega$, $i = 1, 2, \dots, n$ y que satisfacen las siguientes proposiciones:

- $P(E_i) > 0, \forall i = 1, 2, \dots, n$
- $E_i \cap E_j = \phi, \forall i \neq j ; i, j = 1, 2, \dots, n$
- $\bigcup_{i=1}^n E_i = \Omega$

entonces

$$P(A) = \sum_{i=1}^n P(E_i)P(A/E_i) \quad (11)$$

esta expresión es conocida como *regla de las probabilidades totales*.

Además también se puede obtener las probabilidades “a posteriori”, o sea, una vez ocurrido el experimento, saber la probabilidad de que haya ocurrido un evento, sabiendo que este experimento arroja un resultado. Esta nueva visión de probabilidad la da el teorema de Bayes, que se expresa así:

$$P(E_i/A) = \frac{P(E_i)P(A/E_i)}{\sum_{i=1}^n P(E_i)P(A/E_i)} \quad (12)$$

5. Variables Aleatorias

Las variables aleatorias, son funciones que asignan a un evento de un experimento que existe en un espacio muestral, un numero real que lo identifica. Esto mejor dicho es:

$$\begin{aligned} \mathcal{X} : \Omega &\longrightarrow \mathbf{R} \\ A \rightsquigarrow x &= \mathcal{X}(A) \end{aligned}$$

Para estas variables aleatorias (caso discreto), se tiene que cumplir lo siguiente:

- $p(x) \geq 0, \forall x \in \mathbf{A}$, donde $\mathcal{X} = \text{rec}(\mathbf{A})$
- $\sum_{x \in \mathbf{A}} p(x) = 1$

donde $p(x)$ se llama *función de cuantía*. Y de forma análoga para el caso continuo,

- $f(x) \geq 0, \forall x \in \mathbf{A}$, donde $\mathcal{X} = \text{rec}(\mathbf{A})$
- $\int_{-\infty}^{\infty} f(x) = 1$

pero $f(x)$ se conoce como *función de probabilidad o de frecuencia*.

También se define la *distribución de probabilidades* la cual básicamente se define como:

$$F[x] = \sum_{\min(k \in \mathcal{A})}^x p(k)$$

, o

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

y que además cumple las siguientes propiedades:

- $F[x] \leq F[y] ; x \leq y$, o sea cumple la propiedad de ser monótona creciente.
- $\lim_{x \rightarrow -\infty} F[x] = 0$
- $\lim_{x \rightarrow \infty} F[x] = 1$

6. Modelos probabilísticos

Esta sección trata de los modelos probabilísticos básicos, para esto se introduce una nueva notación o especificación, o sea, si tenemos una variable aleatoria X , entonces se puede uno a encontrar que:

$$X \sim \text{nombre}(\text{parametros})$$

esto nos dice que X es variable aleatoria con cierta distribución.

Por ejemplo:

$$X \sim \text{Exp}(\lambda)$$

en este caso X cumple con una distribución exponencial con parámetro λ .

6.1. Modelos discretos

6.1.1. Distribución Bernoulli

Esta distribución es la mas básica de todas, ya que corresponde a cualquier variable aleatoria que asume 1 de 2 estados, y su función de cuantía correspondiente es:

$$\begin{aligned} p(x) &= \mathbf{p}^x (1 - \mathbf{p})^{1-x} \\ x \in \{0, 1\} \quad 0 < \mathbf{p} < 1 \end{aligned} \quad (13)$$

donde \mathbf{p} corresponde a la probabilidad de que el evento ocurra, y $(1 - \mathbf{p})$ es la probabilidad de que no ocurra.

6.1.2. Distribución Binomial

Básicamente se trata de realizar n experiencias Bernoulli, que a su vez son independientes entre si. Además se define el evento X que nos dice el numero de éxitos en los n ensayos. Y por consiguiente la función de cuantía será:

$$\begin{aligned} P(X = x) &= p(x) = \binom{n}{x} \mathbf{p}^x (1 - \mathbf{p})^{n-x} \\ x &= 0, 1, 2, \dots, n \end{aligned} \quad (14)$$

y como es una función de cuantía se cumple que

$$\sum_{x=0}^n \binom{n}{x} \mathbf{p}^x (1 - \mathbf{p})^{n-x} = 1$$

6.1.3. Distribución geométrica

En este caso se define X como el numero de ensayos hasta que aparezca el primer éxito. Entonces

$$\begin{aligned} p(x) &= (1 - \mathbf{p})^{x-1} \mathbf{p} \\ x &= 1, 2, \dots \end{aligned} \quad (15)$$

6.1.4. Distribución binomial negativa

Acá se intenta buscar la probabilidad de obtener el r -ésimo éxito en x ensayos Bernoulli independientes, entonces la función de cuantía correspondiente será la siguiente:

$$\begin{aligned} p(x) &= \binom{x-1}{r-1} \mathbf{p}^r (1 - \mathbf{p})^{x-r} \\ x &= r, r+1, r+2, \dots \end{aligned} \quad (16)$$

6.1.5. Distribución hipergeométrica

Aquí el evento X son el numero de artículos de 1 tipo en n extracciones, pero en las extracciones, con n artículos totales, ocurre sin reposición. Entonces

$$p(x) = \frac{\binom{N_1}{x} \binom{N_2}{n-x}}{\binom{N}{n}} \quad (17)$$

$$\begin{aligned} 0 \leq x \leq N_1 \quad ; \quad n - N_2 \leq x \leq n \\ \max\{0, n - N_2\} \leq x \leq \min\{n, N_1\} \\ N_1 + N_2 = n \end{aligned}$$

6.1.6. Distribución Poisson

Esta distribución se obtiene de extremar la distribución binomial, o sea,

$$\lim_{n \rightarrow \infty, p \rightarrow 0} \binom{n}{x} \mathbf{p}^x (1 - \mathbf{p})^{n-x} = \frac{(n\mathbf{p})^x e^{-n\mathbf{p}}}{x!}$$

donde finalmente la función de cuantía es:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (18)$$

$$x = 0, 1, 2, \dots \quad \lambda = n\mathbf{p}$$

aquí X es el numero de cosas que pasan en un periodo.

6.2. Modelos Continuos

6.2.1. Distribución Uniforme

La principal característica de esta distribución, es que todos los numeros en un intervalo son equiprobables. La densidad viene dada por:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases} \quad (19)$$

Aquí se dice que $X \sim U[a, b]$.

6.2.2. Distribución Normal

Esta distribución es la que se considera mas importante, ya que la mayoría de las medidas que se toman corresponde a este tipo de distribución. La densidad normal correspondiente en forma general es:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (20)$$

Donde μ corresponde a la media o esperanza, además σ corresponde a la desviación estándar o σ^2 es la varianza. Pero lo mas importante es cuando se hace el cambio de variables $Z = \frac{X-\mu}{\sigma}$ ya que cualquier densidad normal se puede dejar expresado en una forma universal y cuya distribución se encuentra tabulada.

$$F[x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad Z = \frac{X-\mu}{\sigma}$$

al hacer el cambio de variable queda expresado por:

$$\Phi(Z) = F\left[Z = \frac{X-\mu}{\sigma}\right] = \int_{-\infty}^Z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Esta distribución esta tabulada con el nombre de *distribución normal estándar*, además como se trabaja con distribuciones es bueno saber que se cumple que:

- $\Phi(0) = 0,5$
- $\Phi(Z) = 1 - \Phi(-Z)$
- $P[c \leq Z \leq d] = \Phi(d) - \Phi(c) \quad c < d$
- $P[Z \geq c] = 1 - P(Z < c) = \Phi(c)$
- $P[Z \geq c] = P(Z \leq -c) = \Phi(-c)$

En este caso la distribución normal se define por 2 parámetros.

$$X \sim \text{Exp}(\mu, \sigma^2)$$

6.2.3. Distribución Gamma

Primeramente es necesario saber que es una función gamma, esta función viene definida por:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

donde esta integral converge y es un numero real positivo, dado $\alpha > 0$.

Alguna de las propiedades de la función gamma son:

- $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha) \quad \forall \alpha \in \mathcal{R}$
- $\Gamma(\alpha + 1) = \alpha! \quad \text{solo si } \alpha \in \mathcal{N}$

Entonces la densidad de la distribución gamma se obtiene de hacer el cambio de variable $x = \lambda y$, resultando en:

$$f(x) = \frac{x^{\alpha-1} \lambda^{\alpha} e^{-\lambda y}}{\Gamma(\alpha)} \quad (21)$$

con

$$\alpha > 0, \quad \lambda > 0, \quad y > 0$$

En este caso se dice que:

$$X \sim \Gamma(\alpha, \lambda)$$

7. Vectores Aleatorios

Cuando se habla de *vectores aleatorios*, se da a entender que son probabilidades o funciones de estas, pero dependen de mas de una variable aleatoria. Pero al igual que antes se cumplen los axiomas:

Si $\underline{X} = (x_1, x_2, \dots, x_n)$ y $f(\underline{X}) = f_{\underline{X}}(x_1, x_2, \dots, x_n)$, entonces:

1. $f_{\underline{X}}(x_1, \dots, x_n) \geq 0 \quad \forall x_i \quad i = \{1, \dots, n\}$
2. $\int_{\underline{X}} f(\underline{X}) = 1$

A esta función $f_{\underline{X}}(\underline{X})$ se le llama *función de densidad conjunta*.

7.1. Distribuciones marginales

Una *distribución marginal* nos da la idea de la forma como depende una probabilidad con respecto a una sola variable. Una *función densidad de probabilidad marginal* se define así:

$$f_X(x) = \int_Y f_{XY}(x, y) dy \quad (22)$$

Aquí hay que tener cuidado, ya que cuando se calcula la densidad conjunta, hay que fijarse bien en el dominio de las otras variables.

7.2. Distribuciones condicionales

Las *distribuciones condicionales* al igual que antes, ver la ecuación (9), viene dada por la intersección de los eventos, esta intersección es igual a la *función de probabilidad conjunta*, entonces ahora las funciones condicionales quedan de la siguiente forma:

$$f_{X_1/X_2}(x_1/x_2) = \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)} \quad (23)$$

8. Transformación de variables aleatorias y vectores aleatorios

No siempre las variables aleatorias que uno quiere, son precisamente las que se pueden medir, en muchas ocasiones uno va tener que transformar la variable medida.

Entonces dado el caso en que tengamos un *vector aleatorio* $\underline{X} = (X_1, X_2)$ y queremos llevarlo a un nuevo vector aleatorio $\underline{Y} = (Y_1, Y_2)$, tal que existe $g(\underline{X}) = \underline{Y}$, es decir, existen

$$g_1(X_1, X_2) = Y_1 \quad \wedge \quad g_2(X_1, X_2) = Y_2$$

y sus correspondientes funciones inversas, o sea

$$g_1^{-1}(Y_1, Y_2) = X_1 \quad \wedge \quad g_2^{-1}(Y_1, Y_2) = X_2$$

de que existan las funciones inversas, implica de que el Jacobiano no se indefina ni se anule, un Jacobiano viene dado por la expresión:

$$|J| = \left| \frac{\partial(g, f)}{\partial(x, y)} \right| = \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix}$$

Si todo lo anterior se cumple entonces, la función cambia de la siguiente forma:

$$\begin{aligned} f_Y(y_1, y_2) &= f_X(g_1^{-1}(Y_1, Y_2), g_2^{-1}(Y_1, Y_2)) \cdot \\ &\quad \cdot \left| \frac{\partial(g_1^{-1}, g_2^{-1})}{\partial(Y_1, Y_2)} \right| \cdot I(Y_1, Y_2) \end{aligned} \quad (24)$$

Cuando se realiza un cambio de variables, hay que tener cuidado ya que los dominios de las nuevas variables suelen cambiar con respecto a las antiguas variables. Para esto se define la *función indicatriz*¹.

Un caso mas sencillo, es el de *Transformación en una variable aleatoria*, simplemente se trata de lo siguiente. Al igual que antes hay ciertas precauciones a tomar, o sea, se tienen que cumplir ciertas condiciones para realizar el cambio, estas eran:

- $g(X) = Y$, pero $g(X)$ es monótona.
- $\frac{d}{dx}(g)$ existe y no se indefina ni se anula en el dominio de x .

entonces la expresión para el cambio de variable es:

$$f_Y(y) = f_X(g^{-1}(Y)) \left| \frac{d}{dy} g^{-1}(Y) \right| \cdot I(y) \quad (25)$$

Aquí también es importante el dominio de la nueva función.

Todo lo visto aquí sobre cambio de variable, es solo valido para el caso continuo, para un caso discreto de cambio de variable, hay que realizar el cambio de variable, yendo “caso por caso”.

9. Distribuciones de varias variables aleatorias

Esta sección tratara de las distribuciones en 2 o más variables aleatorias.

9.1. Distribución Multinomial

La distribución multinomial viene de la generalización de la distribución binomial, que se había visto, cuando se hablaba de una distribución binomial, aparecía el caso de un experimento Bernoulli, o sea, era uno u otro caso, pero

¹Esta función $I(\alpha)$ toma 2 valores, 1 si α pertenece a un dominio ó 0 en otro caso.

para el caso de distribución multinomial ahora aparecen k opciones.

En base de esto la función de cuantía para el caso multinomial es:

$$p_{\underline{X}}(x_1, \dots, x_k) = \frac{n!}{x_1! \dots x_k!} \mathbf{p}_1^{x_1} \dots \mathbf{p}_k^{x_k} \quad (26)$$

$$\text{con } \sum_{i=1}^k x_i = n, \text{ y además } \sum_{i=1}^n \mathbf{p}_i = 1.$$

No es difícil darse cuenta que si $k = 2$ esta distribución multinomial, se vuelve una distribución binomial.

9.2. Distribución Normal Bivariada

Se dice que $\underline{X} = (X_1, X_2)$ es un vector continuo normal bivariado si su densidad está dada por:

$$f_{\underline{X}}(x_1, x_2) = \frac{e^{-\frac{q}{2}}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \begin{pmatrix} I(x_1, x_2) \\ \mathbf{R}^2 \end{pmatrix} \quad (27)$$

donde

$$q = \frac{1}{1-\rho^2} [z_1^2 - 2\rho z_1 z_2 + z_2^2]$$

y además $z_1 = \frac{x_1 - \mu_1}{\sigma_1}$, $z_2 = \frac{x_2 - \mu_2}{\sigma_2}$. Donde los dominios son $(x_1, x_2) \in \mathbf{R}^2$, $\sigma_i \in \mathbf{R}^+$, $\mu_i \in \mathbf{R}$ con $i = 1, 2$ y por último $|\rho| < 1$. Donde más adelante, se verá que significan μ , σ y ρ .

Otra forma de ver la distribución, es escribirla en notación vectorial. Si $\underline{X} = (X_1, X_2)$ y $\underline{\mu} = (\mu_1, \mu_2)$, entonces

$$f_{\underline{X}}(\underline{x}) = \frac{\sqrt{|V^{-1}|}}{2\pi} e^{-\frac{1}{2}(\underline{x} - \underline{\mu})V^{-1}(\underline{x} - \underline{\mu})^T} \begin{pmatrix} I(\underline{x}) \\ \mathbf{R}^2 \end{pmatrix} \quad (28)$$

donde V^{-1} es la matriz inversa de V , que es

$$V = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Para este tipo de distribución tiene la siguiente notación:

$$\underline{X} \sim N_2(\underline{\mu}; V)$$

10. Esperanza o valor esperado

La Esperanza o valor esperado, da la idea de un promedio, más general, al promedio común, ya que este pondera cada valor con su respectiva probabilidad de que ocurra este, de un punto de vista matemático este se define como:

Caso discreto Sea X variable aleatoria discreta, tal que $x \in \mathcal{A}$, y con correspondiente función de cuantía $p(x)$, entonces la esperanza de X , o $E[X]$, se define como

$$E[X] = \sum_{x \in \mathcal{A}} xp(x) \quad (29)$$

$$|E[x]| < \infty \Leftrightarrow \sum_{x \in \mathcal{A}} |x|p(x) < \infty$$

Caso continuo Sea X variable aleatoria continua, tal que $x \in \mathcal{A}$, y con correspondiente función de densidad $f(x)$, entonces la esperanza de X , o $E[X]$, se define como

$$E[X] = \int_{x \in \mathcal{A}} xf(x)dx = \int_{x \in \mathcal{A}} x dF[x] \quad (30)$$

$$|E[x]| < \infty \Leftrightarrow \int_{x \in \mathcal{A}} |x|f(x)dx < \infty$$

10.1. Esperanza de una función

Esto es una generalización de lo anterior, ya que si se utiliza el teorema de transformación de variables (ver la sección 8), se puede llegar a que una esperanza de este tipo es:

$$E[g(X)] = \int_{\mathcal{A}_g} g(x)f(x)dx$$

o bien en el caso discreto

$$E[g(X)] = \sum_{\mathcal{A}_g} g(x)p(x)$$

10.2. Esperanza de un vector aleatorio

Al igual que el caso simple, esto corresponde a un “promedio”, pero esta vez en un espacio. Esta clase de esperanza se define como:

$$E[\underline{X}] = \int_{\mathcal{A}_{\underline{X}}} \underline{X}f_{\underline{X}}(\underline{x})d\underline{x} \quad (31)$$

Aunque esto corresponde a una esperanza con coordenadas ya que:

$$E[\underline{X}] = \int_{\underline{X}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} f(\underline{x})d\underline{x}$$

11. Desviación media, varianza y desviación standard

La desviación media, varianza y desviación standard, miden la dispersion o variabilidad de la variable aleatoria, con respecto a un centro, que puede ser la mediana o la media.

11.1. Desviación media

Básicamente da el valor de dispersion de datos, con respecto a la media, pero se define como la esperanza de la distancia entre media y los datos, es decir,

$$E[|X - \mu|] = \begin{cases} \sum_x |x - \mu| p(x) & X \text{ discreta} \\ \int_x |x - \mu| f(x) dx & X \text{ continua} \end{cases} \quad (32)$$

11.2. Varianza

Al igual que la desviación media, la varianza se define como la esperanza de la distancia entre la media y los valores de los datos, pero esta vez, la medida de distancia es el cuadrado de la diferencia, o sea,

$$Var[X] = E[(X - \mu)^2] \quad (33)$$

pero también se puede anotar la varianza como σ^2 .

11.3. Desviación Standard

Se la desviación standard o σ define como

$$\sigma = \sqrt{E[X^2] - E[X]^2} \quad (34)$$

Esta relación resulta de ocupar las propiedades de la varianza.

12. Desigualdad de Chebyshev

Sea X variable aleatoria con $Var[X] = \sigma^2$, donde $\sigma^2 > 0$ y $E[X] = \mu$. Entonces, para cada $t > 0$ se cumple que:

$$P(|X - \mu| > t) < \frac{\sigma^2}{t^2} \quad (35)$$

13. Covarianza y Correlación

13.1. Covarianza

La idea de covarianza, viene de la necesidad de ver cuan asociadas están dos variables aleatorias. De esta forma se dice que existe una *asociación favorable* si cuando una variable crece la otra también, y de forma contraria, una *asociación desfavorable*, es el caso cuando si una crece, otra variable aleatoria decrece. Se define covarianza, como:

$$\text{cov}(x, y) = E[(X - E[X])(Y - E[Y])] \quad (36)$$

Entonces si $\text{cov}(x, y) < 0$, hay asociación desfavorable, y si $\text{cov}(x, y) > 0$ hay asociación favorable. También cabe recordar que $\text{cov}(x, y)$ solo mide el tipo de asociación, no el grado de asociación. Por eso se puede decir que

$$X, Y \text{ v. a. independientes} \Rightarrow \text{cov}(x, y) = 0$$

Algunas propiedades de la covarianza son:

1. $\text{cov}(x, y) = \int_X \int_Y (x - \mu_x)(y - \mu_y) f_{xy}(x, y) dx dy$
2. $\text{cov}(x, y) = E[xy] - E[x]E[y]$
3. $\text{cov}(x, x) = \text{var}(x)$
4. $\text{cov}(x, y + z) = \text{cov}(x, y) + \text{cov}(x, z)$
5. $\text{cov}(ax, by) = ab \text{ cov}(x, y)$

13.2. Coeficiente o índice de correlación

Si tenemos dos variables aleatorias X, Y tal que $\text{cov}(x, y)$ existe y $\text{var}(x) > 0$, $\text{var}(y) > 0$. Se define la correlación (ρ) entre X e Y como:

$$\rho = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)}\sqrt{\text{var}(y)}} \quad (37)$$

$$-1 \leq \rho \leq 1$$

Notación También se puede denotar el índice de correlación (ρ), las covarianzas y las varianzas como:

$$\begin{aligned} \rho &= \rho_{xy} \\ \text{cov}(x, y) &= \sigma_{xy} \\ \sqrt{\text{var}(x)} &= \sigma_x \\ \sqrt{\text{var}(y)} &= \sigma_y \end{aligned}$$

quedando de una forma mas compacta

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

El resumen el índice de correlación nos da a conocer cuan bien esta relacionada 2 variables aleatorias, si $\rho_{xy} = 1$, quiere decir que existe una relación lineal entre la variables aleatorias.

14. Esperanza y Varianza condicional

El caso de esperanza y varianza condicional, al igual que la probabilidad condicional, estos casos son los mas frecuentes y se pueden llevar al caso no condicional muy fácilmente. La esperanza de una variable dado otra es:

$$E(Y/X) = \begin{cases} \sum_Y y p_{y/x}(y/x) \\ \int_Y y f_{y/x}(y/x) dy \end{cases} \quad (38)$$

Y usando la definición de varianza, resulta que la varianza condicional es:

$$Var[Y/X] = E[Y^2/X] - E^2[Y/X] \quad (39)$$

donde aquí no representa gran dificultad saber cual es la expresión para $E[Y^2/X]$.

Existe el caso, cuando se quiere la esperanza total, partiendo de casos condicionales, esto es:

$$E[Y] = E[E[Y/X]] \quad (40)$$

esto se conoce como *ley de las esperanzas totales*. Aquí hay que especificar que $E[Y/X] = f(X)$ es una función de X , entonces la esperanza total corresponde a:

$$E[Y] = E[f(X)]$$

Y para el caso de varianza, esto se complica ya que la *ley de varianzas totales* es:

$$Var[Y] = Var[E[Y/X]] + E[Var[Y/X]] \quad (41)$$