FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

EYP1113 - Probabilidad y Estadística

Taller R - 06: Distribución conjunta: Caso discreto

Ricardo Aravena C. - Cristian Capetillo C. - Ingrid Guevara R.
Ricardo Olea O. - Bladimir Morales T., Daniel Saavedra M.

Facultad de Matemáticas
Departamento de Estadística
Pontificia Universidad Católica de Chile

Segundo Semestre 2024

Contenido I

Motivación

Distribución conjunta, condicional y marginal

Aplicaciones

Motivación

Las distribuciones conjuntas, marginales y condicionales son herramientas fundamentales en estadística, ya que permiten comprender y **modelar las relaciones** entre múltiples variables aleatorias.

Supongamos que queremos estudiar la relación entre la altura y el sexo en un grupo de personas. Estos dos atributos no son independientes, dado que la altura puede variar en función de otras variables, como el sexo (hombre o mujer).

Para modelar esta relación, utilizamos las distribuciones conjuntas, marginales y condicionales, cada una proporcionando una **perspectiva distinta sobre la interacción** entre las variables.

Distribución conjunta

En el mundo real, muchos fenómenos no dependen solo de una variable aislada, sino de la interacción entre varias.

Si las variables aleatorias X e Y son discretas, la función de distribución de probabilidad **conjunta** es

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

siendo su función de distribución de probabilidad acumulada igual a

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j),$$

con $(x_i, y_j) \in \Theta_{X,Y}$.

Distribución condicional

Es crucial cuando queremos comprender la relación entre variables bajo ciertas condiciones.

Para variables aleatorias discretas X e Y, la probabilidad de $X = x$ puede depender de los valores que puede tomar Y (y viceversa).

En base a lo visto en probabilidades, se define la función de distribución de probabilidad **condicional** como:

$$p_{X|Y=y}(x) = P(X = x \mid Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad p_Y(y) > 0$$

De manera análoga, se tiene que

$$p_{Y|X=x}(y) = P(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}, \quad p_X(X) > 0$$

Distribución marginal

Por último, en muchas situaciones, no nos interesa conocer toda la información disponible. Tal vez solo queramos entender el comportamiento de una variable específica, independientemente de las demás. Esto se describe por la distribución marginal (estrictamente hablando, ya hemos trabajado con ella a lo largo de nuestras clases).

La distribución **marginal** de una variable aleatoria se puede obtener aplicando el teorema de probabilidades totales. Así, la distribución marginal de X en el caso discreto, $p_X(x)$ está dada por

$$p_X(x) = \sum_{y \in \Theta_Y} p_{X,Y}(x,y)$$

De la misma forma se tiene que

$$p_Y(y) = \sum_{x \in \Theta_X} p_{X,Y}(x,y)$$

Aplicación 1: Análisis muestral

Se usará la base de datos de `heightweight` disponible en el paquete `gcookbook`, que contiene información de diferentes características de 236 personas, como la altura, sexo, peso y otros. Se solicita clasificar la altura en tres categorías: bajo (menos de 60 cm), media (entre 60 cm y menos de 65 cm) y alto (65 cm o más).

Aplicación 1: Análisis muestral

Si queremos modelar la probabilidad de observar una combinación específica de altura para un determinado sexo (hombre o mujer) en una persona seleccionada al azar, usamos una distribución **conjunta**.

Supongamos que ya sabemos que una persona tiene estatura baja y queremos determinar la probabilidad de que sea hombre o mujer. Para esto utilizamos la distribución **condicional**.

También podríamos estar interesados únicamente en la altura, sin tener en cuenta el sexo. La distribución **marginal** de la altura nos proporciona información sobre cómo se distribuyen las alturas en la población, sin importar el sexo. De forma similar, podríamos analizar cómo se distribuye el sexo sin tomar en cuenta la altura.

Aplicación 2: Basado en supuestos poblacionales

Se quiere analizar un centro de atención telefónica. Esta información está en el archivo `llamadas.xlsx`. **Supondremos** que el número de llamadas que recibe un centro de ventas en una hora, Y, sigue una distribución de Poisson con un promedio de 6 llamadas por hora, esto es, $Y \sim Poisson(\lambda)$.

Para cada llamada, existe una probabilidad del 60% de que el cliente decida hacer una compra, lo que significa que $X|Y$, el número de compras realizadas dado el número de llamadas recibidas, sigue una distribución binomial con $n = y$ y $p = 0.60$, esto es, $X|Y = y \sim Binomial(y, p)$.

Aplicación 2: Basado en supuestos poblacionales

- ¿Cuál es la probabilidad de que se tenga 5 llamadas por hora y que 4 clientes hicieron una compra?
- ¿Cuál es la probabilidad de se hayan realizado 6 llamadas en una hora?
- ¿Cuál es la probabiliad de que se tengan 2 llamadas por hora dado que no se realizo ninguna compra?