Conceptos básicos de la inferencia estadística

Unidad 3. Repaso de probabilidad (parte 1)

Javier Santibáñez

IIMAS, UNAM

jsantibanez@sigma.iimas.unam.mx

Semestre 2020-1

Contenido

¿Qué es la probabilidad?

Probabilidad (DRAE, 2014)

1. Verosimilitud o fundada apariencia de verdad.
2. Cualidad de probable (que se verificará o sucederá).
3. En un proceso aleatorio, razón entre el número de casos favorables y el número de casos posibles.

Verosímil (DRAE, 2014)

1. Que tiene apariencia de verdadero.
2. Creíble por no ofrecer carácter alguno de falsedad.

¿Qué es la probabilidad?

Aleatorio (DRAE, 2014)

1. Perteneciente o relativo al juego de azar.
2. Que depende del azar.

Azar (DRAE, 2014)

1. Casualidad, caso fortuito.
2. Desgracia imprevista.

Casualidad (DRAE, 2014)

- Combinación de circunstancias que no se pueden prever ni evitar.

Probabilidad

La probabilidad es una medida del grado de creencia que se tiene acerca de que algo incierto ocurra Generalmente se pensará en que se tiene un experimento aleatorio, es decir, del cuál no puede saberse con anticipación cuál será el resultado. A pesar de que no es posible saber el resultado con anticipación, es posible conjeturar sobre la verosimilitud de ocurrir que tiene cada uno de los posibles resultados. El objetivo de la teoría de probabilidad es formalizar esta noción intuitiva de probabilidad.

Espacios de probabilidad

Un espacio de probabilidad se conforma de tres elementos

1. Un conjunto de posibles resultados llamado espacio muestral, denotado por Ω.
2. Una familia de conjuntos que representan los eventos de interés, denotado por $\mathcal{A}$.
3. Una función que asigna las probabilidades a los eventos de $\mathcal{A}$, denotada por P.

La familia de eventos $\mathcal{A}$

La familia de eventos de interés $\mathcal{A}$ debe tener una estructura particular, que permita asignar las probabildiades de manera consistente.

Se dice que $\mathcal{A}$ es una σ-álgebra de elementos de un conjunto Ω si cumple con lo siguiente

1. $\Omega \in \mathcal{A}$.
2. Si $E \in \mathcal{A}$, entonces $E^c \in \mathcal{A}$, donde $E^c = \Omega \setminus E$, es el complemento de E.
3. Si $E_1, E_2, E_3, \ldots \in \mathcal{A}$, entonces

$$\bigcup_{n=1}^{\infty} E_n \in \mathcal{A}.$$

La función P

La función P que asigna las probabilidades a los eventos de interés va de $\mathcal{A}$ al intervalo $[0, 1]$ y es una medida de probabilidad.

Si $\mathcal{A}$ es una σ-álgebra de subconjuntos de un conjunto Ω, se dice que P es una medida de probabilidad en $(\Omega, \mathcal{A})$ si

1. $P(\Omega) = 1$.
2. Para todo $E \in \mathcal{A}$, $P(E) \geq 0$.
3. Si $E_1, E_2, E_3, \ldots \in \mathcal{A}$ y $E_i \cap E_j = \emptyset$, para todos $i \neq j$, entonces

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n).$$

Algunos resultados elementales

Si $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad, entonces

1. $P(E^c) = 1 - P(E)$.
2. Si $E \cap F = \emptyset$, entonces $P(E \cup F) = P(E) + P(F)$.

Espacios equiprobables

Algunos de los ejemplos más sencillos de espacios de probabilidad son los espacios en los que todos los resultados individuales en Ω tienen la misma probabilidad. Por ejemplo, en el lanzamiento de un dado.

En los espacios equiprobables las probabilidades de los eventos se pueden obtener como el cociente del número de casos favorables entre el total de resulados posibles.

Por lo anterior, las técnicas de conteo son relevantes en los cursos básicos de probabilidad.

Principios básicos de conteo

Principio de adición

Si una tarea se puede llevar a cabo de n o m maneras distintas, entonces la tarea se puede realizar de $m + n$ formas diferentes.

Principio del producto

Este principio establece que si una operación se puede hacer de n formas y cada una de estas pueden llevarse a cabo m maneras distintas en una segunda operación, se dice que juntas las operaciones pueden realizar se de $n \times m$ formas distintas.

Ejemplos

Juegos de cartas

Si se selecciona al azar una carta de una baraja inglesa, ¿cuál es la probabilidad de que se obtenga un trébol?

Arreglos

Cinco parejas se sentarán al azar en una mesa redonda, ¿cuál es la probabilidad de que ninguna pareja se siente junta?

Juegos de dados

¿Cuál es la probabilidad de sacar como suma un 7 cuando se lanzan dos dados?

Probabilidad condicional

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, si $E, F \in \mathcal{A}$ y $P(F) > 0.$, se define la probabilidad condicional de E dado F como

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Resultados importantes

Teorema de probabilidad total

Si $E_1, E_2 \in \mathcal{A}$ son tales que $E_1 \cup E_2 = \Omega$ y $E_1 \cap E_2 \emptyset$, entonces entonces

$$P(F) = P(F \mid E_1)P(E_1) + P(F \mid E_2)P(E_2), \qquad \text{para todo } F \in \mathcal{A}$$

Teorema de Bayes

Si $E, F \in \mathcal{A}$ y $P(E), P(F) > 0$, entonces

$$P(F|E) = \frac{P(E \mid F)P(F)}{P(E)}$$

Independencia

El concepto de independencia es fundamental en estadística, ya que proporciona las bases para la teoría de estimación.

Eventos independientes

Sean $E, F \in \mathcal{A}$. Se dice que E y F son independientes si

$$P(E \cap F) = P(E)P(F)$$

Si $P(E), P(F) > 0$ la definición anterior implica que

$$P(E \mid F) = E \qquad \text{y} \qquad P(F \mid E) = P(F)$$

Variables aleatorias

En ocasiones no interesa sino estudiar características numéricas de los resultados del experimento. Esta noción se formaliza con la definición de variable aleatoria.

Se dice que una función $X : \Omega \to \mathbb{R}$ es una variable aleatoria si para todo $a \in \mathbb{R}$ se cumple que

$$X^{-1}(-\infty, a] \in \mathcal{A}.$$

donde X^{-1} denota la imagen inversa de X

$$X^{-1}(-\infty, a] = \{\omega \in \Omega : X(\omega) \leq a\}.$$

Función de distribución acumulada

Si X es una variable aleatoria definida en un espacio de probabilidad $\Omega, \mathcal{A}, P)$, se define la función de distribución de X como la función $F : \mathbb{R} \to [0, 1]$ tal que

$$F(x) = P(X \leq x), \qquad -\infty < x < \infty.$$

Las funciones de distribución acumulada cumplen con las siguientes propiedades

1. $\lim_{x \to -\infty} F(x) = 0$ y $\lim_{x \to \infty} F(x) = 1$.
2. F es continua por la derecha: $\lim_{x \to a^+} F(x) = F(a)$.
3. F es no decreciente: si $x_1 < x_2$, entonces $F(x_1) \leq F(x_2)$.

Tipos de variables aleatorias

El rango de una variable aleatoria se define como el conjunto de posibles valores de la variable y se denota por $X(\Omega)$

$$X(\Omega) = \{X(\omega) : \omega \in \Omega\}$$

Las v.a. se clasifican según su rango.

1. Si el rango de una v.a. es un conjunto numerable, se dice que la v.a. es discreta.
2. Si el rango de una v.a. es un conjunto no numerable, casi siempre un intervalo de números reales, se dice que la v.a. es continua.

Variables aleatorias discretas

Las variables aleatorias discretas están caracterizadas por su función masa de probabilidades (fmp). Si X es una v.a. discreta, se define la fmp de X como la función $p_X : X(\Omega) \to [0, 1]$ tal que

$$p_X(x) = P(X = x), \qquad x \in X(\Omega).$$

La fmp de una v.a. discreta cumple con la siguiente propiedad

$$\sum_{x \in X(\Omega)} p_X(x) = 1.$$

Momentos de una v.a. discreta

Si X es una v.a. discreta con fmp p_X, se define el valor esperado o esperanza de X como

$$E(X) = \sum_{x \in X(\Omega)} x p_X(x),$$

siempre que la suma sea finita. Se define la varianza de X como

$$V(X) = \sum_{x \in X(\Omega)} (x - E(X))^2 p_X(x)$$

Variables aleatorias continuas

Las variables aleatorias continuas están caracterizadas por su función de densidad de probabilidad (fdp). Si X es una v.a. la fdp de X es la función $f_X : \mathbb{R} \to [0, \infty)$ y es tal que

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Las fdp de v.a. continuas cumplen con la siguientes propiedades

$$\int_{X(\Omega)} f_X(x)dx = 1 \qquad \text{y} \qquad \frac{d}{dx} F_X(x) \bigg|_{x=t} = f_X(t).$$

Momentos de una v.a. discreta

Si X es una v.a. continua con fdp f_X, se define el valor esperado o esperanza de X como

$$E(X) = \int_{X(\Omega)} x f_X(x) dx$$

siempre que la integral sea finita. Se define la varianza de X como

$$V(X) = \int_{X(\Omega)} (x - E(X))^2 f_X(x) dx.$$

Momentos de variables aleatorias

Dada una variable aleatoria X y una función $g : \mathbb{R} \to \mathbb{R}$, se cumple que

$$E(g(X)) = \int_{X(\Omega)} g(x) f_X(x) dx$$

Dada una variable aleatoria X, se define el k-ésimo momento no central de X, $k = 1, 2, 3, \ldots$ como $E(X^k)$ y se define el k-ésimo momento central de X como $E((X - \mu)^k)$, donde $\mu = E(X)$, para $k = 1, 2, 3, \ldots$.

Función generadora de momentos

La función generadora de momentos de una v.a. X se define como

$$M_X(t) = E(e^{tX}).$$

Esta función es útil por el siguiente resultado

$$\frac{d^k}{dx^k} M_X(t) \bigg|_{t=0} = E(X^k), \qquad k = 1, 2, 3, \ldots.$$

Distribuciones más comunes en estadística

Distribuciones discretas

- Distribuciones de Bernoulli y binomial
- Distribuciones geométrica y binomial negativa
- Distribución hipergeométrica
- Distribución de Poisson

Distribuciones continuas

- Distribución uniforme
- Distribución exponencial
- Distribución normal
- Distribuciones gamma y beta.

Distribuciones de Bernoulli y binomial

Un ensayo de Bernoulli es un experimento aleatorio con solamente dos posibles resultados: 0 (fracaso) y 1 (éxito).

Se dice que una v.a. X tiene una distribución binomial si X cuenta el número de éxitos obtenidos en una serie de ensayos de Bernoulli idénticos e independientes. Los parámetros de la distribución son: el número de ensayos n y la probabilidad de éxito en cada ensayo $\theta, \theta \in (0, 1)$.

La fmp de una v.a. X con distribución $Bin(n, \theta)$ es

$$p_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

La distribución de Bernoulli es la distribución $Bin(1, \theta)$.

Distribuciones geométrica y binomial negativa

El problema inverso (negativo) a contar el número de éxitos obtenidos en n ensayos es contar el número de ensayos necesarios para obtener r éxitos.

Se dice que una v.a. X tiene una distribución binomial negativa si cuenta el número de ensayos de Bernoulli idénticos e independientes necesarios para obtener cierto número de éxitos. Los parámetros de esta distribución son: el número de éxitos requeridos r y la probabilidad de éxito en cada ensayo.

La fmp de una v.a. X con distribución $BinNeg(r, \theta)$ es

$$p_X(x) = \binom{x-1}{r-1}\theta^r(1-\theta)^{x-r}, \qquad x = r, r+1, r+2, \ldots .$$

La distribución geométrica corresponde a la distribución $BinNeg(1, \theta)$.

Distribución hipergeométrica

Considerar una urna con N bolas en total, de las cuales n son rojas y el resto son negras. Suponer que se extraen al azar m bolas de la urna. Una variable tiene una distribución hipergeométrica si cuenta el número de bolas rojas extraídas. Los parámetros de la distribución son: el total de bolas en la urna$N$, el número de bolas rojas b y el número de bolas que se extraen m. La fmp de una v.a. X con distribución $HiperG(N, n, m)$ es

$$p_X(x) = \frac{\binom{n}{x}\binom{N-n}{m-x}}{\binom{N}{m}}, \qquad \text{máx}(0, n - N + m) \leq x \leq \text{mín}(n, m)$$

Distribución de Poisson

La distribución de Poisson se utiliza para modelar procesos que ocurren en el tiempo. Por ejemplo, el número de terremotos que ocurren en un mes, el número de accidentes viales que se producen en una hora, el número de reclamaciones que se reciben en una semana, entre otras. El parámetro de la distribución es λ, $\lambda > 0$, la tasa de eventos por unidad de tiempo.

Se dice que una v.a. X tiene una distribución $Poi(\lambda)$ si su fmp está dada por

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots.$$

Distribución uniforme

La distribución uniforme asigna probabilidades iguales a los intervalos de la misma longitud. Esta distribución se utiliza para generalizar el enunciado "seleccionado al azar" para el caso continuo.

Se dice que una v.a. X tiene una distribución $U(a, b)$, con $a, b \in \mathbb{R}$ y $a < b$, si su fdp está dada por

$$f_X(a) = \frac{1}{b-a}\mathbb{1}_{(a,b)}(x), \qquad -\infty < x < \infty,$$

donde $\mathbb{1}_{(a,b)}(x)$ es la función indicadora del intervalo (a, b).

Distribución exponencial

La distribución exponencial se utiliza para modelar tiempos de vida, principalmente. Esta distribución es la única que tiene la propiedad de perdida de memoria.

Se dice que una v.a. continua X tiene una distribución $Exp(\lambda)$, $\lambda > 0$, si su fdp está dada por

$$f_X(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) \mathbb{1}_{(0,\infty)}(x), \qquad -\infty < x < \infty.$$

En algunos casos se utiliza otra parametrización de la distribución $Exp(\lambda)$, pero en el curso utilizaremos esta.

Distribución normal

Esta es quizá la distribución de probabilidades más conocida de todas. También recibe el nombre de distribución gaussiana. Esta distribución se utilizó originalmente para modelar la distribución de errores de medición.

Se dice que una v.a. continua X tiene una distribución $N(\mu, \sigma^2)$, si su fdp está dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right), \qquad -\infty < x < \infty.$$

Funciones gamma y beta

Función gamma

Se define la función gamma como la función $\Gamma : (0, \infty) \to (0, \infty)$ tal que

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

Función beta

Se define la función beta como la función $B : (0, \infty)^2 \to (0, \infty)$ tal que

$$B(x, y) = \int_0^1 t^{x-1} (1 - t)^{y-1} dt$$

Propiedades de las funciones Γ y B

1. La función Γ es una generalización del factorial

$$\Gamma(n) = (n-1)!, \qquad n = 1, 2, 3, \ldots .$$

2. Las funciones Γ y B se relacionan como sigue

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

3. La función B es una función simétrica

$$B(x,y) = B(y,x).$$

4. Se puede demostrar que

$$\Gamma(0.5) = \sqrt{\pi}.$$

Distribuciones gamma y beta

Se dice que una v.a. continua X tiene una distribución $Ga(\alpha, \lambda)$, $\alpha, \lambda > 0$, si su fdp está dada por

$$f_X(x) = \frac{x^{\alpha-1}}{\lambda^\alpha \Gamma(\alpha)} \exp\left(-\frac{x}{\lambda}\right) \mathbb{1}_{(0,\infty)}(x), \qquad -\infty < x < \infty.$$

Se dice que una v.a. continua X tiene una distribución $Be(\alpha, \beta)$, $\alpha, \beta > 0$, si su fdp está dada por

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbb{1}_{(0,1)}(x), \qquad -\infty < x < \infty.$$

Distribuciones derivadas de la normal

Las siguientes distribuciones surgen cuando se realizan inferencias en poblaciones normales

- Distribución χ^2
- Distribución t
- Distribución F.

En esta sección se presentan algunos resultados importantes relacionados con estas distribuciones.

Distribución χ^2

La distribución χ^2 aparece en muchos partes en estadística, principalmente en pruebas de hipótesis.

Se dice que una v.a. continua X tiene una distribución χ^2 con parámetro ν, $\nu > 0$, lo cual se denota como $X \sim \chi^2_{(\nu)}$, si su fdp está dada por

$$f_X(x) = \frac{x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \exp\left(-\frac{x}{2}\right) \mathbb{1}_{(0,\infty)}(x), \qquad -\infty < x < \infty.$$

Típicamente al parámetro ν se le nombra *grados de libertad*.

Esta distribución se relaciona con la normal a través del siguiente resultado.

Si $Z_1, \ldots, Z_n$ son variables aleatorias $N(0,1)$ independientes, entonces

$$Z_1^2 + \ldots + Z_n^2 \sim \chi^2_{(n)}$$

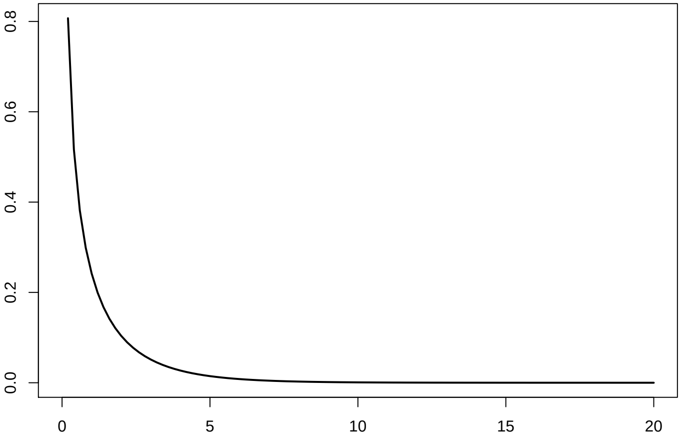

Figura: fdp de la distribución $\chi^2_{(\nu)}$.

Distribución χ^2

La distribución χ^2 es un caso particular de la distribución gamma

$$\chi^2_{(\nu)} = Ga\left(\frac{\nu}{2}, \frac{1}{2}\right)$$

Este hecho permite obtener algunos resultados sobre la distribución χ^2

Si $X \sim \chi^2_{(\nu)}$, $\nu > 0$, entonces

1. $E(X) = \nu$.
2. $V(X) = 2\nu$.

Distribución t

La distribución t fue propuesta por William Gosset en 1908, cuando hacía análisis de calidad en Guinness Brewery. Por motivos de confidencialidad tuvo que publicarla bajo el pseudónimo *student*.

Esta distribución surge cuando se hacen inferencias sobre la media de una distribución normal y se asume que la varianza es desconocida.

Se dice que una v.a. continua X tiene una distribución t con parámetro ν, $\nu > 0$, lo cual se denota como $X \sim t_{(\nu)}$, si su fdp está dada por

$$f_X(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad , -\infty < x < \infty.$$

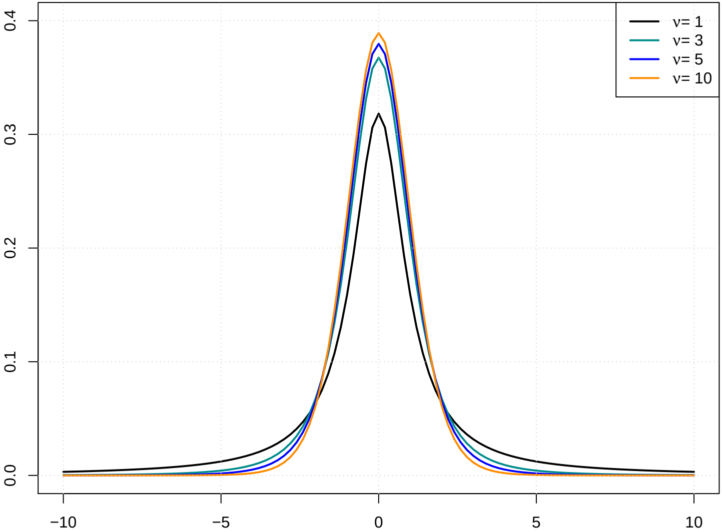

Figura: fdp de la distribución $t_{(\nu)}$.

Distribución t

El siguiente resultado muestra la relación entre la distibución t y la distribución normal.

Si $X \sim N(0,1)$ y $Y \sim \chi^2_{(\nu)}$ son variables aleatorias independientes, entonces

$$U = \frac{X}{\sqrt{Y/\nu}} \sim t_{(\nu)}.$$

Además, si $f(x \,|\, \nu)$ denota la densidad $t_{(\nu)}$ y $\phi(x)$ denota la fdp $N(0,1)$, entonces

$$\lim_{\nu \to \infty} f(x \,|\, \nu) = \phi(x), \qquad \text{para todo } x \in \mathbb{R}.$$

Distribución F

La distribución F surge del Análisis de Varianza.

Se dice que una variable aleatoria continua X tiene una distribución F con parámetros ν_1 y ν_2, $\nu_1, \nu_2 > 0$, lo que se denota como $X \sim F_{(\nu_1, \nu_2)}$, si su fdp se está dada por

$$f_X(x) = B^{-1}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1}\left(1 + \frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1+\nu_2}{2}} \mathbb{1}_{(0,\infty)}(x),$$

$-\infty < x < \infty$.

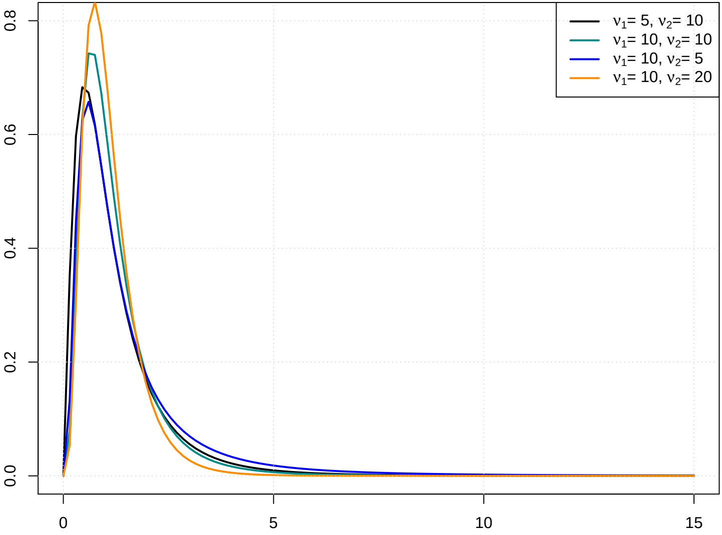

Figura: fdp de la distribución $F_{(\nu_1,\nu_2)}$.

Distribución F

Si $X \sim \chi_{(\nu_1)}$ y $Y \sim \chi_{(\nu_2)}$ son variables aleatorias independientes, entonces

$$Z = \frac{X/\nu_1}{Y/\nu_2} \sim F_{(\nu_1,\nu_2)}.$$

Se puede mostrar que la si $X \sim F_{(\nu_1,\nu_2)}$, entonces $Y = 1/X \sim F_{(\nu_2,\nu_1)}$. Además

$$E(X) = \frac{\nu_2}{\nu_2 - 2}, \qquad \text{si } \nu_2 > 2.$$

Cuantiles distribucionales

Para responder a la pregunta ¿cuál es la probabilidad de que una v.a. X sea menor o igual a q? se usa la fda, ya que

$$F_X(c) = P(X \leq c)$$

El planteamiento inverso es el siguiente ¿cuál es la constante c tal que la probabilidad de que X sea mayor a q sea igual a α? esto es, ¿cual es el valor de q tal que

$$F_X(q) = P(X \leq q) = \alpha?$$

En el primer caso se fija q y se calcula α, mientras que en el segundo se fija α y se busca q.

La función de cuantiles

Cuantiles distribucionales y función de cuantiles

Sea X una v.a. con fda F_X. Al mínimo número q tal que $F_X(q) \geq \alpha$ se le nombra cuantil α de la v.a. X, para cualquier α en $(0, 1)$. Generalmente se denota a los cuantiles como q_p o $q_{(p)}$.

Se define la función de cuantiles de X como la función $Q : (0, 1) \to \mathbb{R}$ tal que

$$Q(\alpha) = \text{mín}\left\{q : F_X(q) \geq \alpha\right\}, \qquad \alpha \in (0, 1).$$

- Los cuantiles de algunas distribuciones tienen notaciones especiales: normal estándar (z), χ^2, t y F.
- Para las distribuciones continuas $Q = F^{-1}$.