Variables Aleatorias (III). Distribuciones habituales.

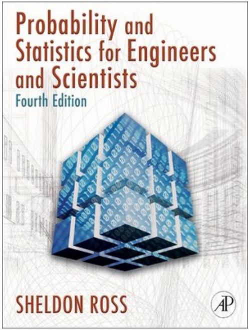
Estadística, Grado en Sistemas de Información

Constantino Antonio García Martínez

Universidad San Pablo Ceu

Contenidos

References



S. Ross. Introduction to Probability and Statistics for Engineers and Scientists. Chapter 5.

C.D. Barr, D.M. Diez, M. Çetinkaya-Rundel. OpenIntro Statistics. Chapters 3.

Distribuciones discretas

Distribuciones discretas

Distribución binomial

Distribución binomial

Ejercicio: Distribución binomial
Se tira una moneda n veces. Si p es la probabilidad de cara, ¿cómo se distribuye la VA X: número de caras?

Jacob Bernouilli.

Distribución binomial
En un experimento que repetimos varias veces, llamamos a cada repetición **ensayo**. Estamos interesados en contar el número de **éxitos** X, donde un éxito ocurre con probabilidad p y no ocurre con probabilidad $q = 1 - p$. Si los ensayos son independientes entre sí (**ensayos de Bernouilli**) obtenemos la distribución binomial:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \cdots, n.$$

Denotaremos $X \sim \mathcal{B}(n, p)$. El caso $n = 1$ suele llamarse **distribución de Bernouilli**

Ejemplo: Probabilidad Binomial

La probabilidad de obtener 2 caras en 6 lanzamientos es

$$P(X=2)=\binom{6}{2}0.5^2(0.5)^4=\frac{15}{64}.$$

Ejercicio: Esperanzas

Expresa X como $X=I_1+I_2+\cdots+I_n$, donde I_i es la función indicatriz del suceso "Éxito" en el lanzamiento i-ésimo, para demostrar que:

$$\mu=np \qquad \sigma^2=npq$$

Ejemplo: Esperanzas

En 100 lanzamientos de moneda con $p=1/3$ el número esperado de caras es $\mu=100/3\approx 33.3$ mientras que la desviación típica es $\sigma=\sqrt{100\frac{1}{3}\frac{2}{3}}\approx= 4.71$.

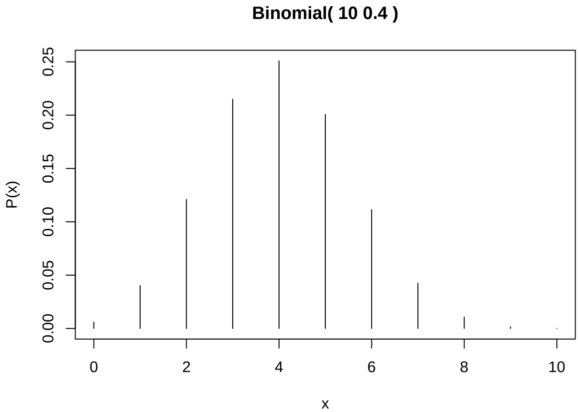
Distribución Binomial en R

R tiene comandos para trabajar con las distribuciones más famosas. Estos comandos comienzan con *d*, *p*, y *r*. Luego toman un sufijo que describe la distribución. Para la distribución Binomial:

Comando	¿Qué hace?
dbinom(k,n,p)	$P(X = k)$
pbinom(k,n,p)	$P(X \leq k)$
qbinom(probs,n,p)	Calcula los cuantiles especificados por las *probs*
rbinom(k,n,p)	Simula k VAs

Ejemplo: Distribución de Bernouilli

```
n = 10
p = 0.4
x = seq(0, 10, 1)
plot(x, dbinom(x, n, p), xlab = "x", ylab = "P(x)",
     main = paste("Binomial(", n, p, ")"), type = "h")
```



Distribuciones discretas

Distribuciones geométrica, hipergeométrica y Binomial negativa

Existen muchas distribuciones que responden a preguntas interesantes acerca de experimentos de Bernouilli.

Ejemplo:
Se tira una moneda con probabilidad de cara p ¿Cuál es la probabilidad de que se obtenga la primera de ellas en la tirada x-ésima?

Existen muchas distribuciones que responden a preguntas interesantes acerca de experimentos de Bernouilli.

Ejemplo:
Se tira una moneda con probabilidad de cara p ¿Cuál es la probabilidad de que se obtenga la primera de ellas en la tirada x-ésima?

Distribución geométrica
La distribución geométrica cuenta el número de fallos X antes de obtener el primer éxito en una serie de experimentos de Bernouilli.

$$P(X = x) = (1 - p)^x p.$$

Escribimos $X \sim \mathrm{Geom}(p)$. La esperanza y los momentos son:

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{1-p}{p^2}.$$

En `R` tenemos las funciones *dgeom*, *pgeom*, ...

Danger!
A veces la distribución geométrica se define como el número de intentos X hasta obtener el primer éxito. En tal caso:

$$P(X = x) = (1 - p)^{x-1} p.$$

En este curso daremos prioridad a la otra parametrización por ser la que usa `R`.

Distribución binomial negativa

La distribución geométrica cuenta el número de fallos X hasta obtener el primer éxito en una serie de experimentos de Bernouilli. La distribución binomial negativa cuenta el número de fallos hasta el r-ésimo éxito.

Ejemplo:
Se tira una moneda con probabilidad de cara p. ¿Cuáles la probabilidad de que se obtenga la quinta cara en la novena tirada?

Distribución binomial negativa
Cuenta el nº de fallos X hasta el r-ésimo éxito.

$$P(X=x) = \binom{r+x-1}{r-1} p^r q^x \qquad x = 0, 1, 2, \ldots$$

Escribimos $X \sim \text{NegBin}(r, p)$. La esperanza y los momentos son:

$$\mu = \frac{rq}{p} \qquad \sigma^2 = \frac{rq}{p^2}.$$

En R tenemos los comandos *dnbinom*, *pnbinom*, *rnbinom*, ...

Distribución hipergeométrica

La distribución binomial aparece en el muestreo con reemplazamiento, mientras que la **distribución hipergeométrica aparece en el muestreo sin reemplazamiento.**

Ejemplo:
Tenemos una caja con m bolas blancas y n bolas rojas. Extraemos k de ellas sin reemplazamiento. ¿Cuál es la probabilidad de obtener x bolas blancas?

Sea X el número de éxitos (bolas blancas en el ejemplo anterior) en k ensayos:

$$P(X = x) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}} \qquad x = 0, 1, 2, \cdots$$

La media y la varianza son:

$$\mu = \frac{km}{m+n} = kp \qquad \sigma^2 = \left(\frac{m+n-k}{m+n-1}\right) k\frac{\mu}{k}(1 - \frac{\mu}{k})$$

donde $p = \frac{m}{m+n}$, $q = \frac{n}{m+n}$ y $N = m + n$:

En R tenemos los comandos *dhyper*, *phyper*, *rhyper*, ...

Distribuciones discretas

Distribución multinomial

Distribución multinomial
Los eventos A_1, $A_2 \cdots A_k$ son mutuamente excluyentes y ocurren con probabilidades p_1, p_2, p_k (verificándose $\sum_{i=1}^{k} p_i = 1$). Si $X_1, X_2, \cdots, X_k$ son variables aleatorias que cuentan el número de veces que los eventos A_1, $A_2 \cdots A_k$ ocurren en n ensayos (Con $\sum_{i=1}^{k} = X_i = n$). Entonces

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

Las esperanzas son

$$\mathbb{E}[X_i] = np_i \qquad \mathrm{Var}[X_i] = np_i(1 - p_i).$$

Los comandos en `R` siguen la nomenclatura esperada: *rmultinom*, *dmultinom*, ...

Ejemplo: Distribución multinomial
Se tira un dado 12 veces. ¿Cuál es la probabilidad de obtener dos 1s, dos 2s, ... y dos 6s?

Distribución multinomial

Los eventos A_1, $A_2 \cdots A_k$ son mutuamente excluyentes y ocurren con probabilidades p_1, p_2, p_k (verificándose $\sum_{i=1}^{k} p_i = 1$). Si $X_1, X_2, \cdots, X_k$ son variables aleatorias que cuentan el número de veces que los eventos A_1, $A_2 \cdots A_k$ ocurren en n ensayos (Con $\sum_{i=1}^{k} = X_i = n$). Entonces

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k) = \frac{n!}{x_1!x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

Las esperanzas son

$$\mathbb{E}[X_i] = np_i \qquad \mathrm{Var}[X_i] = np_i(1 - p_i).$$

Los comandos en `R` siguen la nomenclatura esperada: *rmultinom*, *dmultinom*, ...

Ejemplo: Distribución multinomial

Se tira un dado 12 veces. ¿Cuál es la probabilidad de obtener dos 1s, dos 2s, ... y dos 6s?

$$P(X_1 = 2, X_2 = 2, \cdots, X_6 = 2) = \frac{6!}{2!2! \cdots 2!}(1/6)^2(1/6)^2 \cdots (1/6)^2 \approx 0.0034.$$

Distribuciones discretas

Distribución de Poisson

Distribución de Poisson

Una VA binomial/multinomial tiene prefijado el número máximo de casos de éxito. ¿Qué ocurre si n no está acotado?

Ejemplo:
VAs discretas no acotadas

- El número de bebés nacidos en un hospital en una semana.
- El número de soldados muertos por una patada de caballo en un cuerpo de la caballería prusiana.
- El número de partículas emitidas en la desintegración de una partícula alpha.
- El número de accidentes en una carretera.

Siméon Denis Poisson.

Distribución de Poisson

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \cdots$$

Denotaremos $X \sim \mathcal{P}(\lambda)$.

Ejercicio: Media y Varianza

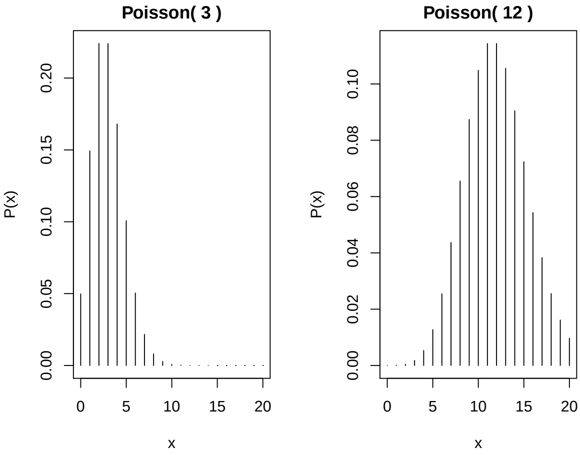
Demuestra que la la media y varianza de la distribución de Poisson son

$$\mu = \lambda \qquad\qquad \sigma^2 = \lambda.$$

Distribución de Poisson en `R`

Comando	¿Qué hace?
dpois(k, lambda)	$P(X = k)$
ppois(k, lambda)	$P(X \leq k)$
qpois(probs, lambda)	Calcula los cuantiles especificados por las *probs*
rpois(k, lambda)	Simula k VAs

Ejemplo: Distribución de Poisson

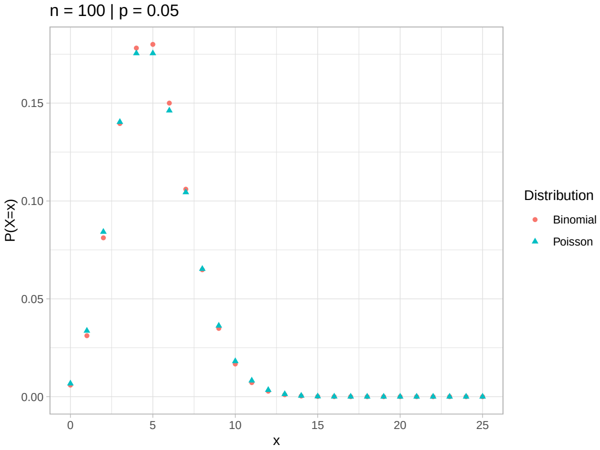


Aproximación de una distribución Binomial mediante Poisson
En una distribución Binomial, si n es grande y p es pequeño, los eventos modelados serán **eventos raros**. En la práctica consideraremos que n es grande y p pequeño si:

$$n \geq 50 \qquad y \qquad np < 5.$$

En tal caso podemos aproximar la distribución Binomial $\mathcal{B}(x \mid n, p)$ con la distribución $\mathcal{P}(\lambda = np)$.

Ejercicio:
Aproximación de una distribución Binomial mediante Poisson Verifica está aproximación usando las funciones de distribución Binomial y de Poisson en R.



Distribuciones continuas

Distribuciones continuas

La distribución uniforme

La distribución uniforme

La VA X tiene distribución uniforme si su función de densidad es

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases},$$

Denotamos $X \sim \mathcal{U}(a, b)$. La media y varianza son,

$$\mu = \frac{1}{2}(a+b) \qquad \sigma^2 = \frac{1}{12}(b-a)^2.$$

Los comandos en `R` son *runif*, *dunif*, ...

Distribuciones continuas

La distribución exponencial

Distribución exponencial

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

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

$$\mu = \frac{1}{\lambda} \qquad \sigma^2 = \frac{1}{\lambda^2}$$

En R tenemos los comandos *dexp*, *pexp*, *rexp*, ... Una propiedad fundamental de la distribución exponencial es que **no tiene memoria**:

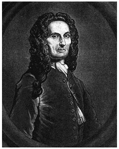
$$P(X < t_1 + t_2 \mid X > t_1) = P(X < t_2).$$

Esta propiedad hace que un uso habitual de la distribución exponencial sea modelar tiempos de espera, tiempos hasta el primer fallo, etc.

Distribuciones continuas

Distribución normal

La distribución normal o Gaussiana es, posiblemente, la más importante de todas las distribuciones.




De Moivre y Gauss.

Distribución Normal
La variable aleatoria X tiene distribución normal con media μ y varianza σ^2 si la función de densidad de X es

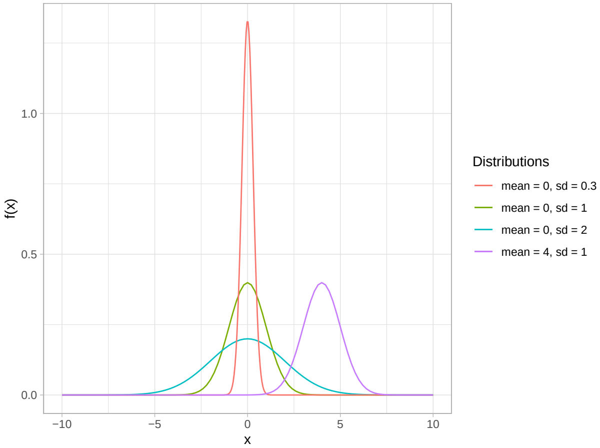
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Escribimos $X \sim \mathcal{N}(\mu, \sigma^2)$.

En R tenemos las funciones *dnorm*(x, μ, σ), *pnorm*(x, μ, σ), *rnorm*(x, μ, σ).

Ejercicio:
Prueba que la media y la varianza de $\mathcal{N}(\mu, \sigma^2)$ son μ y σ^2.

Distribución normal



Distribución normal estandarizada
Sea $X \sim \mathcal{N}(\mu, \sigma^2)$. Definimos la variable normal estandarizada como

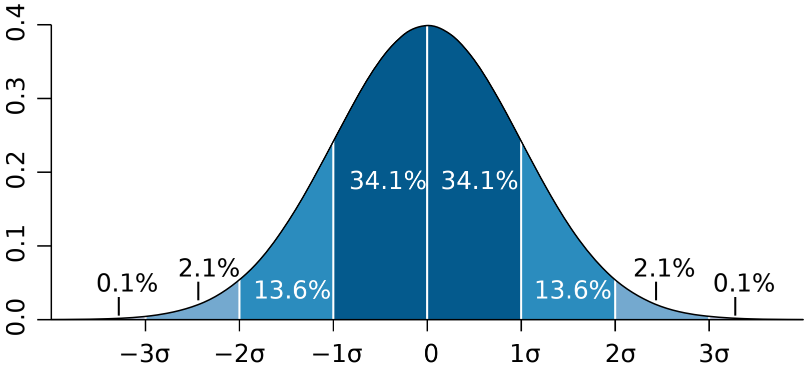
$$Z = \frac{X - \mu}{\sigma}.$$

Ejercicio:
Demuestra que la variable aleatoria Z tiene media 0 y varianza 1.

Para calcular cualquier probabilidad relacionada con X, podemos emplear las tablas de la distribución normal estandarizada.

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P(Z \leq \frac{x - \mu}{\sigma}).$$

Ejemplo:
Demuestra la **regla de 68-95-99**:



Si n es grande y ni p ni q son muy pequeñas, la distribución de la variable aleatoria binomial X binomial puede aproximarse usando que

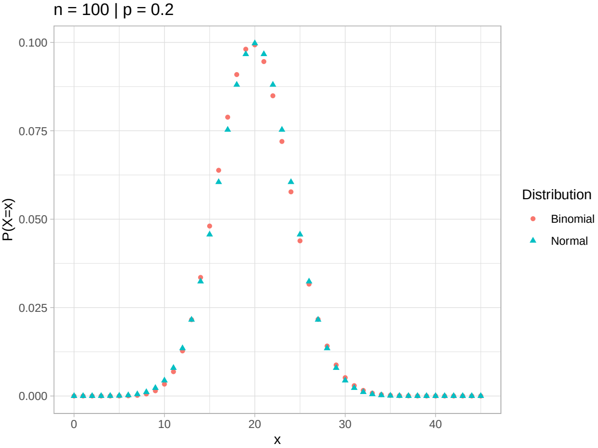
$$Z = \frac{X - np}{\sqrt{npq}}$$

es asintóticamente normal. En la práctica, la aproximación es buena si

$$np > 5 \qquad \text{y} \qquad nq > 5.$$

Ejercicio:
Verifica esta aproximación usando las funciones de distribución Binomial y Normal en `R`.

Aproximación normal de las distribuciones binomiales

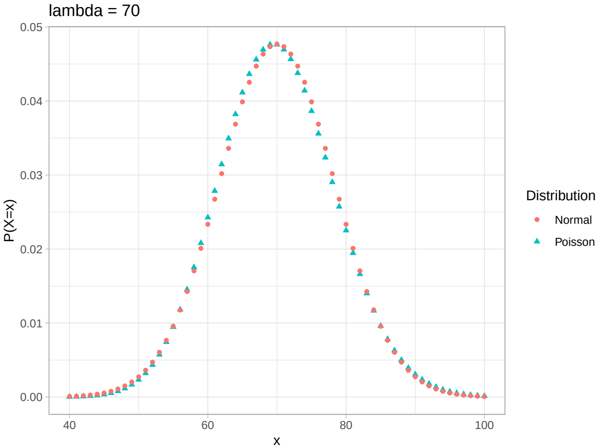


Aproximación normal de las distribuciones Poisson

Si X es una VA de Poisson, la VA

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

se aproxima a una normal a medida que $\lambda \to \infty$.



La similitud entre las aproximaciones Binomial y de Poisson mediante una Gaussiana nos lleva a pensar si existen más distribuciones que tiendan a una distribución Normal.

Teorema del límite central
Si X_1, $X_2, \cdots, X_n$ son VAs independientes e identicamente distribuidas con media μ y varianza σ^2 finitas. Entonces $S_n = \sum_{i=1}^{n} X_n$ verifica

$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \right) = F_z(b) - F_z(a).$$

Es decir, la VA $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ es asintóticamente normal.

Distribuciones continuas

Distribuciones derivadas de la Normal

Distribución de una suma de Normales Independientes

La distribución Normal verifica la **Propiedad reproductiva** [1]. Es decir, si X_1 y X_2 son VAs Normales independientes:

$$X_1 + X2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

[1] Otras VAs que verifican la propiedad reproductiva son la Binomial y la distribución de Poisson

Distribución Chi-cuadrado

Si $X_1, X_2, \cdots, X_\nu$ son ν variables **normales estandarizadas independientes** entonces $\chi^2 = \sum_{i=1}^{\nu} X_i^2$ tiene distibución Chi-cuadrado. ν recibe el nombre de **grados de libertad**. Escribemos $X \sim \chi^2(\nu)$ o $X \sim \chi^2_\nu$ Su media y varianza son:

$$\mu = \nu \qquad \sigma^2 = 2\nu.$$

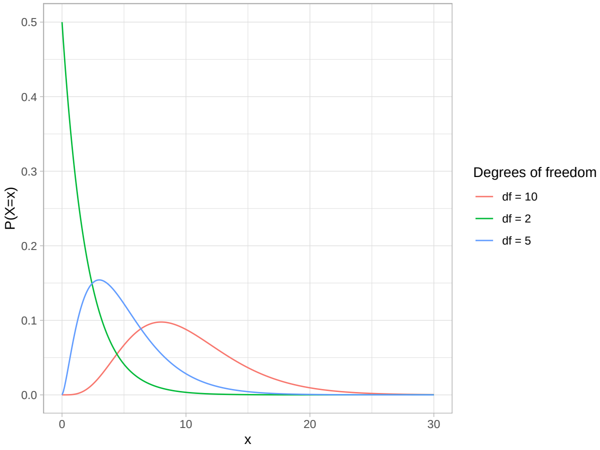
Las funciones en `R` son *dchisq*, *rchisq*, ...

Teorema
Sean $U_1, U_2, \cdots, U_k$ son VAs aleatorias independientes con distribución Chi-cuadrado con grados de libertad $\nu_1, \nu_2, \cdots, \nu_k$. Entonces su suma

$$W = \sum_{i=1}^{k} U_i$$

es una Chi-Cuadrado con $\nu = \sum_{i=1}^{k} \nu_i$ grados de libertad.

Distribución Chi-cuadrado



Gosset.

Si Z y χ^2_ν son VAs independientes, siendo Z una VA normal estandarizada y χ^2_ν una Chi-cuadrado con ν grados de libertad, entonces la VA

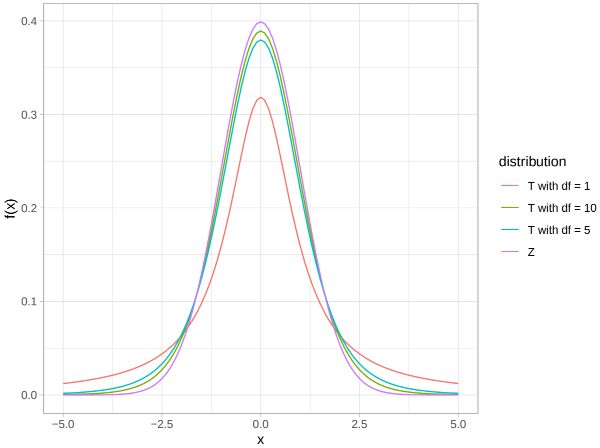
$$T_\nu = \frac{Z}{\sqrt{\chi^2_\nu/\nu}}$$

tiene una distribución T-Student con ν grados de libertad. Escribemos $X \sim t(\nu)$ o $X \sim t_\nu$. Su media y varianza son:

$$\mu = 0 \qquad \sigma^2 = \frac{\nu}{\nu - 2} \text{ si } \nu > 2.$$

Las funciones en `R` son *dt*, *rt*, ...

Distribución T-Student



Distribución F

Fisher y Snedecor.

Sean χ_n^2, χ_m^2 dos VAs Chi-cuadrado independientes con n y m grados de libertad. Entonces la VA

$$F_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}$$

distribución F, F de Snedecor o de Fisher-Snedecor con n, m grados de libertad.

Escribemos $X \sim F(n,m)$ o $X \sim F_{n,m}$. Su media y varianza son:

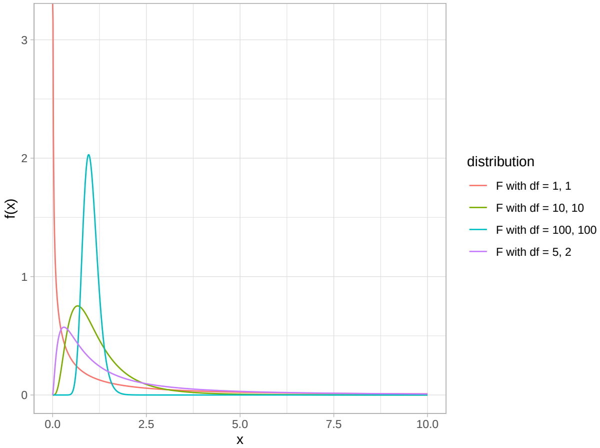
$$\mu = \frac{m}{m-2} \; m > 2 \qquad \sigma^2 = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2} \; m > 4.$$

Las funciones en `R` son *df*, *rf*, ...

Teorema: Cuantiles de F

$$F_{p;n,m} = \frac{1}{F_{1-p;m,n}}$$

Distribución F



Ejercicios

Ejercicio:
En un laboratorio hay 100 ratas completamente aisladas las unas de las otras. Cada rata tiene un 10 % de probabilidad de infectarse por un parásito. ¿Cuál es la probabilidad de que más de 5 se infecten?

Ejercicio:
Los bebés pesan de media 3.4 Kg al nacer con una desviación típica de 0.6 Kg. ¿Cuál es la probabilidad de que un bebé pese más de 4 Kg al nacer?

Ejercicio:
Tiramos un dado 12 veces, cuál es la probabilidad de obtener dos 1s, dos 2s, dos 3s, ... y dos 6s.

Ejercicio:
En cierto hospital, un 12 % de los pacientes no acude a su cita. Si un equipo médico es capaz de atender 100 personas en un día, cuál es la probabilidad de que una persona se quede sin atender si se dan 110 citas en un día. ¿Cuántas citas se pueden dar sin que dicha probabilidad exceda el 5 %?

Ejercicio:
De 2000 familias con 4 niños cada una, cuántos te esperarías que tuviesen a) al menos 1 niño, b) 2 niñas, c) 1 o 2 niñas, d) ningún niño.

Ejercicio:
Has programado un robot asesino para acabar con tu profesor de estadística. El robot dispara al centro de su frío corazón, pero comete un error aproximadamente normal en cada una de las coordenadas x e y. La media de ambas normales es 0 y tiene desviación típica 5 cm. ¿Cuál es la probabilidad de que el disparo acabe a menos de 1 cm del centro del corazón? (No temas por el profesor, es un cyborg indestructible venido del futuro).

Ejercicios

Ejercicio:
Hay una infección en la población de un campus de 1000 personas. Se desea estimar la cantidad de infectados a partir de una muestra de 50 alumnos. El muestreo es sin reemplazamiento y da como resultado que hay 5 infectados. ¿Cómo estimamos la cantidad de infectados en el campus?

Ejercicio:
La media de peso de 500 estudiantes es 68.5 Kg con desviación estándar 6.8 Kg. a) ¿Cuántos estudiantes pesarán entre 54 y 70 Kg? b) ¿Cuántos estudiantes pesarán más de 84 Kg?

Ejercicio:
Un enfermero necesita 10 radiografías de la pierna de un niño. Hay un 70 % de probabilidad de que el niño esté quieto durante la prueba. ¿Cuál es la probabilidad de que se necesiten más de 15 pruebas?

Ejercicio:
Datos de un hospital de maternidad sugieren que nacen 4.5 bebés cada día. ¿Cuál es la probabilidad de que haya 6 partos mañana?

Ejercicio:
Un genetista desea estimar el nº de mutaciones y cross-overs en el proceso de recombinación. Un lab mide, para 100 muestras, los siguientes datos:

nº crossovers	0	1	2	3
nº muestras	50	25	20	5

¿Probabilidad de que en una nueva muestra aparezcan al menos 1 cross-over?

Ejercicio:
Tres hermanos van a comer a casa de su abuelita. El 99.7 % de las veces, cada hermano es capaz de comer una cantidad de comida comprendida entre 1.2 y 1.8 Kg. ¿Cuál es la probabilidad de que logren acabar los 5 Kg de carne que su abuela ha preparado?

Ejercicio:
Se ha estimado que la prob. de mutación por nucleótido es de $1.5 \cdot 10^{-8}$ por nucleótido. Hay $3.3 \cdot 10^{9}$ nucleótidos en el genoma humano. ¿Cuál es la probabilidad de que 80 nucleótidos muten? Usa distintas aproximaciones para contestar a la pregunta.