DISCRETAS
CONTINUAS
MIXTAS

TEMA 3: VARIABLES ALEATORIAS II

Ensayos de Bernouilli → Discretas

Distribución Binomial

- Cada repetición → "Ensayos"
- Ensayos Indep entre sí ⇒ solo hay 2 resultados < Éxito / Fracaso

$\underbrace{\qquad\qquad\qquad\qquad}$

Ensayos de Bernouilli

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?

Distribución binomial
La VA X: "nº de éxitos en n-ensayos de Bernouilli" tiene **distribución binomial**:

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

El caso $n=1$ suele llamarse **distribución de Bernouilli.**

En R:

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

$\mu = n \cdot p$

$\sigma = n \cdot p \cdot q$
↓
$(1-p)$

Distribución Geométrica

Cuenta el nº de FALLOS antes de obtener el 1^er^ éxito

$$X \sim \mathrm{Geom}(p) \iff P(X=x) = \mathrm{Geom}(x \mid p) = (1-p)^x p.$$

Esperanza: $\mu = \frac{q}{p}$ Varianza: $\sigma^2 = \frac{q}{p^2}$ $q = 1-p$

Ejercicio:
Se tira una moneda hasta que sale la primera cara. Si la probabilidad de cara es p, ¿cuál es la distribución de la VA X: "nº de cruces hasta la primera cara"?

dgeom, rgeom, ...

Ejercicios:

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?

X = "nº pruebas hasta 10 radiografías bien"
≃ Binomial Negativa ⇒ $X \sim \mathrm{NegBin}(10, 0.7)$

Y = "nº fallos hasta alcanzar 10 radiogr. bien"

Pregunta → $P(Y > 15) \rightsquigarrow Y = X + 10$
→ $P(X+10 > 10) = P(X > 5)$
↓
Programar en R:

$P(X>5) = 1 - P(X<5) = 1 - F(5)$

└--→ 1 - pnegbinom (5, $\underbrace{\text{size}}_{\text{nº de casos éxito}}$ = 10, Prob=0.7)

Distribución Binomial Negativa

Cuenta el nº Fallos hasta el r-ésimo éxito

$$X \sim \mathrm{NegBin}(r,p)$$

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

Esperanza: $\mu = \frac{r \cdot q}{p}$ Varianza: $\sigma^2 = \frac{r \cdot q}{p^2}$

Ejercicio:
Se tira una moneda con probabilidad de cara p. Sea la VA X: "nº de cruces hasta obtener 5 caras". ¿Cuál es la distribución de esta variable aleatoria?

dnbinom, pnbinom, rnbinom, ...

Ejercicio:
De 2000 familias con 4 niños cada una, cuántos te esperarías que tuviesen a) al menos 1 niño (masculino) y b) 2 niñas.

X: "nº ♂ en una familia"
Y: "nº fam con al menos 1 niño"

a) $Y \sim B(\underbrace{2000}_{\text{tope}}, \underbrace{P(X \geq 1)}_{\text{Prob éxito}})$

$E[Y] = 2000 \cdot P(X \geq 1)$ → $B(\underbrace{4}_{\text{4 niños}}, 0'5)$ └→ suponemos probabilidad

$= 1 - P(X<0) = 1 - P(X=0) = 1 - \text{dbinom}(0, 4, 0'5)$

$1 - P(X<1)$ XX

b) 2 niños

X = "nº niñas por fam"
Y = "nº fam con dos niñas"

resolver con simulaciones

```
#X~B(4, 0.5)
#Y~B(2000, P(X=2))
rbinom

  #P(X=2) => por simulación
  rbinom(1, size= 4, prob = 0.5) #nº simulaciones(1)
  N=5000
  n_girls = rbinom(N, size= 4, prob = 0.5)
  p_success = sum(n_girls == 2) / N

  #E[Y]
  mean(
  rbinom(N, size=2000, p_success)
  )
```

TEMA 3: VARIABLES ALEATORIAS II

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 %?

X: "nº Pacientes q acude a su cita"

Pregunta ▸ P("♀ se quede sin atender") = P(X > 100)

1. Asumiendo Indep ⇒ $X \sim B(110, 0'88)$ $\xrightarrow{}$ 100-12%

$P(X>100) = 1 - P(X \leq 100) = 1 - \underline{F(100)}$

pbinom (100, 110, 0.88)

b) $X \sim B(n, 0.88)$ ⇒ $P(X>100) \leq 0.05$

¿?

$1 - \text{Pbinom}(n, 110, 0.88) \leq 0.05$

en R ⇒ Podemos coger un vector e ir probando

$\boxed{100:110}$ pq con 100 ♀ las atiende a todas 100%

```
n = 100:110
  binomial = 1- pbinom(100, n, 0.88)
  binomial <= 0.05
  tabla = table(binomial <= 0.05)
  tabla[2]+99
```

⇒ TRUE
108

Código R

Distribución Hipergeométrica

Muestreo sin reemplazo

De una urna con m bolas blancas y n bolas negras extraemos k bolas sin reemplazamiento. Sea X: "número de bolas blancas extraídas". Entonces:

$$X \sim \text{HypGeom}(m, n, k)$$

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

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

dhyper, phyper, rhyper, ...

Distribución Multinomial

Ensayos indep pero cada ensayo múltiples resultados

Ejercicio:
Se tira un dado 12 veces. ¿Cuál es la probabilidad de obtener dos 1s, dos 2s, ... y dos 6s? ¿Y la distribución general para x_1 1s, x_2 2s, ...?

Distribución multinomial
Sea $\boldsymbol{X} = [X_1, X_2, \cdots, X_k]$ el vector que cuenta el número de éxitos para k-categorías excluyentes en n ensayos independientes. Si en cada ensayo las probabilidades de éxito para cada categoría son $\boldsymbol{p} = [p_1, p_2, ..., p_k]$, entonces:

$$\boldsymbol{X} \sim \text{MultiNom}(n, \boldsymbol{p}) \iff P(\boldsymbol{X} = \boldsymbol{x}) = \text{MultiNom}(\boldsymbol{x} \mid n, \boldsymbol{p}) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

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

Distribuciones de Poisson

¿Qué ocurre si una variable no está acotada? < nº accidentes en una carretera / bebés nacidos en una semana

• Eventos < independientes / raros (raro q ocurra a la = 0) ⇒ Poisson

Prob. de q ocurra un det. suceso un nº de k veces en un intervalo det. de tiempo ▸ a partir de la f de ocurrencia media → Wikipedia

• Distribución de Poisson

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

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

$\mu = \lambda$

$\sigma^2 = \lambda$

dpois, ppois, ...

Ejercicio:
Un informático desea modelar el nº ataques informáticos que su página web sufre cada día. Recolecta datos durante 100 días, obteniendo la siguiente **tabla de frecuencias**:

nº de ataques/día	0	1	2	3
frecuencia	45	35	15	5

¿Probabilidad de que, en un día cualquiera, haya al menos 1 ataque informático?

no sabes cuál es el tope
• binneg
• Geom
• Poiss → podemos sacar fácilmente nº medio ataques

X = "nº ataques en un día" → $P(X \geq 1)$

$X \sim P(\lambda)$

$\lambda = \text{Media} = \frac{35 \cdot 1 + 15 \cdot 2 + 5 \cdot 5}{100} = 0.8$

Dibujar representación:
xs: 0:6 valores x
Plot (xs, dpois(xs, 0.8))

$P(X \geq 1) = 1 - P(X<0) = 1 - P(X=0) = 1 - \text{dpois}(0, 0.8) = 0.5$

Ejercicio:
Se tira una moneda hasta que salen 10 caras. A continuación, se meten 10 bolas blancas y tantas bolas negras como cruces hayan salido en una urna. Se extraen 5 bolas de la urna sin reemplazamiento. ¿Cuál es el número más probable de bolas blancas y cuál es su probabilidad?

nº cruces = nº caras - 10 = nº bolas negras

X = "nº cruces hasta 10 caras"

Y = "nº bolas blancas extraídas"

Pregunta ⇒ nº q más veces va a salir de Y ⇒ $\underline{\underline{\text{Moda}}}$

Es necesario calcular P(Y) para todos los y's ⇓ Calculo f(x) de Prob

→ $X \sim \text{NegBin}(10, 0.5)$

$Y \mid X \sim \text{HypGeom}(10, X, 5)$ ↗ bolas q saco; b blancas, b negras

$P(Y = y's)$:

$P(Y \mid X) = \frac{P(Y, X)}{P(X)}$

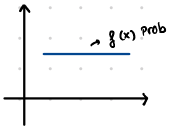
1. Calculamos $P(Y, X)$
2. Marginalizamos

TEMA 3: VARIABLES ALEATORIAS II

Distribuciones continuas

Distribución uniforme

Todos los puntos son equiprobables

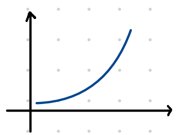


$X \sim U(a,b) \Leftrightarrow f(x) = U(x|a,b) \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{e.o.c.} \end{cases}$

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

runif, dunif, ...

Distribución exponencial



$X \sim Expo(\lambda) \Leftrightarrow f(x) = Expo(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$

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

dexp, pexp, rexp, ...

tiempos de espera, tiempos hasta 1er fallo

````
Se está monitorizando la confiabilidad de tres ordenadores de un clúster. Sean
$X_a$, $X_b$ y $X_c$ los tiempos que transcurren hasta que cada uno de estos
ordenadores se bloquean y necesitan reiniciarse. Estos "tiempos de bloqueo"
son independientes entre sí y tienen distribuciones exponenciales con parámetros
respectivos $\lambda_a = 2, \lambda_b = 3$ y $\lambda_c = 5$. Sea $M$ el instante
de tiempo en que el primero de estos ordenadores falla (¡puede ser cualquier de los
tres!). Verifica mediante simulaciones que la distribución de M tiene
distribución exponencial de parámetro 10. Para ello:

a. Escribe el código necesario para simular la variable $M$. Guarda el resultado
de tus simulaciones en la variable `time_sims`. (Pista: las funciones
`pmin` o `pmax` pueden ser útiles.)

Datos:
Xa, Xb, Xc independientes -> a=2, b= 3, c=5
```{r block_i}

time_sims = function(n){
  p_a = rexp(n,2)
  p_b = rexp(n,3)
  p_c = rexp(n,5)
  pmin(p_a, p_b, p_c)
}

b. Estima la distribución a partir de las simulaciones usando un **histograma**
   (función `hist`; consulta la ayuda para dibujar una densidad de probabilidad
   en lugar del número de ocurrencias) y...
c. ... compara el histograma obtenido con la distribución teórica (la
   exponencial de parámetro 10). Para ello dibuja en un mismo gráfico la
   distribución obtenida  por simulaciones y la distribución teórica
   (usando `dexp` y `lines`).

```{r block_ii_and_iii}
#b)
hist(time_sims(5000), freq = FALSE)
#c)
x = seq(0,1, by = 0.01)
lines(x,dexp(x, 10), type = "l", col = 2)
````

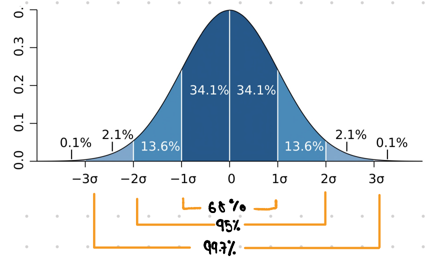
Distrib. Normal / de Gauss

$X \sim N(\mu, \sigma^2)$

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

dnorm (x, mean, sd), pnorm (q, mean, sd), ...

⚠ sd mean= 4 → "Centrada entorno a 4"



Teorema del Límite Central:

Suma muchas VA's indep → tiende a parecerse a una normal

$$S_n = \sum_{i=1}^{n} X_n \Longrightarrow S_n \Rightarrow N(n\mu, n\sigma^2)$$

* Si el nº de variables sumadas es pequeño → No funciona

Propiedad reproductiva de Normales Indep

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Ejercicio:
Tres hermanos van a comer a casa de su abuelita. El 99.7 % de las veces, cada hermano come 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?

X = "kg carne comido" → $X = X_1 + X_2 + X_3$

X_1 = "kg carne comido â 1"

X_2 = "kg carne comido â 2"

X_3 = "kg carne comido â 3"

$P(X>5)$

como es continua

$P(X>5) = P(X \ge 5)$

Lo más razonable es que más o menos coman lo mismo todos los días

$X_i \sim N(\mu, \sigma)$

$\mu = \frac{1'2 + 1'8}{2} = 1'5$

$\sigma \rightarrow -3\sigma \quad 3\sigma \quad 99'7\% \quad \rightsquigarrow \quad \begin{matrix} 1'5 + 3\sigma = 1'8 \\ 1'5 - 3\sigma = 1'2 \end{matrix} \Big\} \sigma = 0.1$

Asumimos indep ⇒ Suma 3 variables indep q son Gaussianas:

$E[X] = E[X_1+X_2+X_3] = E[X_1] + E[X_2] + E[X_3]$

$= 1'5 + 1'5 + 1'5 = 4'5$

$Var[X] = Var[X_1] + Var[X_2] + Var[X_3]$

$= (0.1)^2 + (0.1)^2 + (0.1)^2 = 0.03$

$\sigma = \sqrt{0.03} = 0.1\sqrt{3}$

$X \sim N(4'5, 0.1\sqrt{3}) \Rightarrow P(X>5) = 1 - \underbrace{\underbrace{P(X<5)}_{F(5)}}_{\text{pnorm}}$

TEMA 3: VARIABLES ALEATORIAS II

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? Resuelve por simulaciones.

D = "distancia al cx del Corazón"
X = "error en eje x" ⇝ $X \sim U(\mu=0, \sigma=5cm)$
Y = "error en eje Y" ⇝ $Y \sim U(\mu=0, \sigma=5cm)$

¿Cómo se distribuye D? ⇒ SIMULACIONES

```
# sd = σ

N=500
x=rnorm(N, 0, sd=5)
y=rnorm(N, 0, sd=5)
d= sqrt(x^2 + y^2)
sum(d<1)/N
```

→ d<1 ▸ caso éxito

Distribuciones Mixtas

Ejercicio:
Trabajamos para una gran compañía de streaming que quiere hacer sugerencias de películas a sus usuarios. Para hacer esto, hay que intentar caracterizar los gustos de cada usuario. Asumimos que las películas que ve un usuario únicamente dependen de dos rasgos de su caracter: su "aventurosidad (X)" y su "romanticismo (Y)". Asumimos que ambos rasgos se pueden caracterizar con $X \sim \mathbb{U}[0,1]$ y $Y \sim \mathbb{U}[0,1]$ (independientes). Si un usuario tiene como rasgos $X = x$ e $Y = y$:

- La probabilidad de ver una película de acción es $p_a = \frac{2x-y+1}{6}$.
- La probabilidad de ver una película romántica es $p_r = \frac{3y-2x+2}{10}$
- La probabilidad de ver una película de acción-amor es $p_{ar} = 1 - p_a - p_r$.

De las últimas 20 películas que ha visto un usuario, 4 han sido de acción, 4 románticas y 12 de acción-amor. ¿Cuáles son sus rasgos X e Y más probables?

A: "nº películas acción"
R: "nº películas románticas"
M: "nº películas acción-amor"

$\boxed{\text{discreta}}$ ¿tiene tope? sí: 20 películas → Binomial / Hypergeom / multinomial

$(A,R,M) \sim \text{Multinomial}(20, [P_a, P_r, P_m])$

dependen de x e y

$(A,R,M \mid x,y) \sim \text{Multinomial}(20, [P_a, P_r, P_m])$