Esenciais USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

DISTRIBUCIÓNS DISCRETAS
funcións notables, momentos e representacións gráficas

María Isabel Borrajo García, Mercedes Conde Amboage e Rosa María Crujeiras Casais (Departamento de Estatística, Análise Matemática e Optimización)

O mundo real está cheo de fenómenos aleatorios que son gobernados pola incerteza, nos que non coñecemos con veracidade o seu resultado. Este tipo de fenómenos formalízanse matematicamente a través dos denominados experimentos aleatorios: probas que, repetidas en idénticas condicións, poden dar lugar a resultados diferentes.

A cuestión é que non sempre nos interesa o resultado en si mesmo dun experimento aleatorio, senón un valor numérico determinado polo seu resultado. A correspondencia que permite transformar os resultados dun experimento aleatorio nunha cantidade numérica é o que se coñece como **variable aleatoria**.

Exemplo 1: consideremos o experimento aleatorio consistente en lanzar 10 moedas ao aire e anotar se sae cara, C, ou cruz, $+$. Os posibles resultados deste experimento son coleccións de 10 elementos, no que cada un deles pode ser C ou $+$. No canto de empregar un vector que resulta pouco manexable, podemos resumir a súa información a través dunha cantidade numérica, por exemplo, o número de caras. Definindo a variable aleatoria X =“número de caras en 10 lanzamentos”, se temos que $X = 6$, xa sabemos que o número de cruces será 4.

As variables aleatorias permiten establecer unha correspondencia entre os posibles resultados dun experimento aleatorio e unha cantidade numérica. Dependendo da tipoloxía das posibles asignacións numéricas asociadas, distinguimos entre **variables aleatorias discretas**, aquelas que toman unha cantidade finita ou infinita numerable de valores, e **variables aleatorias continuas**, aquelas que toman valores nun intervalo ou unión de intervalos da recta real.

Función de masa de probabilidade e función de distribución

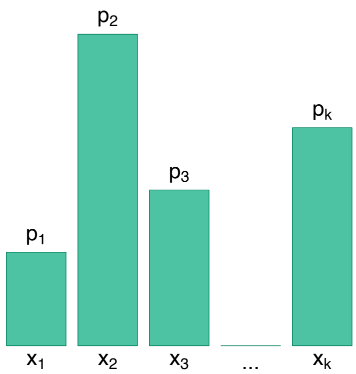
Sexa X unha variable aleatoria discreta que toma os valores $\{x_1, x_2, \ldots, x_k\}$, definimos a **función de masa de probabilidade** como as probabilidades $\{p_1, \ldots, p_k\}$ verificando

$$\mathbb{P}(X = x_i) = p_i, \text{ con } i = 1, \ldots, k, \text{ onde } p_i \geq 0 \text{ e } \sum_{i=1}^{k} p_i = 1.$$

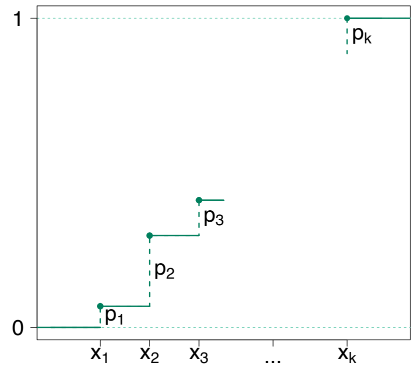
A **función de distribución** dunha variable aleatoria X, F_X, defínese como a función que asocia a cada valor $x \in \mathbb{R}$, a probabilidade de que a variable tome valores menores ou iguais a el, isto é,

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{\substack{x_i \leq x \\ i=1,\ldots,k}} p_i.$$

A función de distribución é unha función non decrecente, que toma valores en $[0, 1]$, e que cumpre que $\lim_{x\to-\infty} F_X(x) = 0$ e $\lim_{x\to+\infty} F_X(x) = 1$. Para as variables aleatorias discretas, a función de distribución é unha función descontinua con saltos nos puntos x_i de altura p_i, para $i = 1, \ldots, k$. De xeito ilustrativo, represéntase a función de masa de probabilidade e a función de distribución asociada a unha variable discreta con valores $\{x_1, x_2, \ldots, x_k\}$ e probabilidades $\{p_1, \ldots, p_k\}$:



Función de masa de probabilidade



Función de distribución

Momentos ou medidas características

A media, a varianza e a desviación típica son medidas características asociadas a calquera variable aleatoria, e ás variables aleatorias discretas en particular. Definimos estas medidas a continuación.

Media

Dada unha variable aleatoria discreta X, con valores $\{x_1, x_2, \ldots, x_k\}$ e probabilidades $\{p_1, p_2, \ldots, p_k\}$, a **media** ou **esperanza poboacional** de X defínese como

$$\mu_X = \mathbb{E}\left(X\right) = \sum_{i=1}^{k} x_i p_i.$$

Varianza

Dada unha variable aleatoria discreta X, con valores $\{x_1, x_2, \ldots, x_k\}$ e probabilidades $\{p_1, p_2, \ldots, p_k\}$, definimos a **varianza poboacional** de X como

$$\sigma_X^2 = \mathbb{V}ar\left(X\right) = \sum_{i=1}^{k} (x_i - \mu_X)^2 p_i.$$

Desviación típica

A **desviación típica poboacional** defínese como a raíz cadrada da varianza:

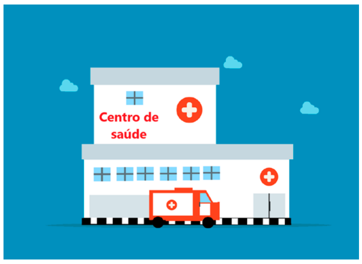
$$\sigma_X = \sqrt{\mathbb{V}ar(X)} = \sqrt{\sum_{i=1}^{k} (x_i - \mu_X)^2 p_i}.$$

A desviación típica presenta unha importante vantaxe fronte á varianza, pois mídese nas mesmas unidades que a variable aleatoria de interese, polo que os seus valores son máis intuitivos e máis facilmente interpretables, mentres que a varianza se mide en unidades cadráticas.

Uniforme discreta

Un caso particular de distribución discreta é aquela na que todas as probabilidades toman o mesmo valor, e coñécese como distribución uniforme discreta. Se supoñemos que a variable ten k posibles valores, a probabilidade asociada a cada un dos valores será $1/k$.

Exemplo 2: consideremos a variable aleatoria $X =$"Número de doentes no servizo de urxencias por hora" dun certo centro sanitario de atención primaria. Sabemos que esta variable aleatoria toma 5 valores diferentes, do 0 ao 4, e a súa función de masa de probabilidade vén dada por:



x_i	0	1	2	3	4
p_i	0.1	0.2	0.3	0.3	0.1

A función de distribución calcúlase de xeito acumulativo a partir da función de masa de probabilidade como segue:

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ 0.1, & \text{se } 0 \leq x < 1, \\ 0.3, & \text{se } 1 \leq x < 2, \\ 0.6, & \text{se } 2 \leq x < 3, \\ 0.9, & \text{se } 3 \leq x < 4, \\ 1, & \text{se } x \geq 4. \end{cases}$$

Para o cálculo da media aplicamos a definición correspondente:

$$\mu_X = \sum_{i=1}^{5} x_i p_i = (0 \cdot 0.1 + 1 \cdot 0.2 + \cdots + 4 \cdot 0.1) = 2.1 \text{ doentes}.$$

De xeito análogo obtemos a varianza e a desviación típica, onde precisamos incluír o valor da media xa obtido:

$$\sigma_X^2 = \sum_{i=1}^{5} (x_i - \mu_X)^2 p_i = \left((0 - 2.1)^2 \cdot 0.1 + \cdots + (4 - 2.1)^2 \cdot 0.1\right) = 1.29 \text{ doentes}^2 \quad \text{e} \quad \sigma_X = \sqrt{1.29} = 1.14 \text{ doentes}.$$

Distribucións notables discretas

Bernoulli

Consideremos un experimento que presenta só dous posibles resultados (por exemplo, éxito ou fracaso, sá/san ou enferma/o, ...). En xeral, denominamos éxito, E, á ocorrencia do evento de interese, e fracaso, F, a que non ocorra. Ademais, a probabilidade de éxito é constante, e denotámola por p. Este tipo de experimentos denomínanse experimentos Bernoulli.

Sexa X a variable aleatoria que rexistra o resultado do experimento, dise que esa variable segue unha distribución de Bernoulli de parámetro p, e denótase como $X \in Ber(p)$:

$$X = \begin{cases} 1, & \text{se ocorre E,} \\ 0, & \text{se ocorre F (non ocorre E).} \end{cases}$$

A función de masa de probabilidade é:

x_i	0	1
p_i	$1-p$	p

e a función de distribución vén dada por:

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ 1-p, & \text{se } 0 \leq x < 1, \\ 1, & \text{se } x \geq 1. \end{cases}$$

A partir da función de distribución podemos calcular a media ou esperanza, e a varianza:

$$\mu_X = \mathbb{E}(X) = p \qquad \text{e} \qquad \sigma_X^2 = \mathbb{V}ar(X) = p(1-p).$$

Binomial

Podemos repetir m veces nas mesmas condicións un experimento Bernoulli, e considerar a variable aleatoria X que conta o número de éxitos obtidos nas m repeticións. Esta variable segue unha distribución binomial de parámetros m e p, e denótase como $X \in Bi(m,p)$.

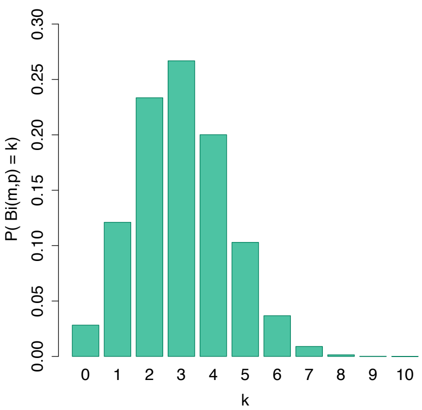
A distribución binomial pode tomar os seguintes valores $\{0, 1, 2, 3, \ldots, m\}$, cuxas probabilidades son:

$$\mathbb{P}(X = k) = \binom{m}{k} p^k (1-p)^{m-k}, \text{ con } k = 0, 1, 2, 3, \ldots, m, \quad \text{sendo } \binom{m}{k} = \frac{m!}{k!(m-k)!}.$$

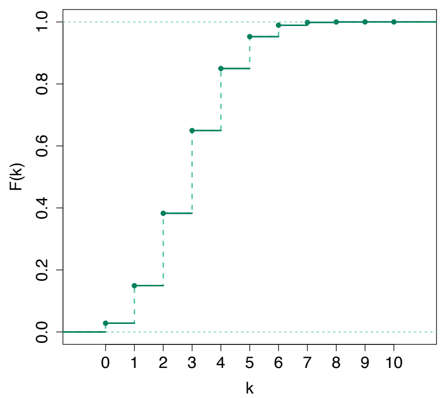
Na distribución binomial, a media e a varianza son:

$$\mu_X = mp \quad \text{e} \quad \sigma_X^2 = mp(1-p).$$

Ilustramos a continuación a función de masa de probabilidade e a función de distribución asociadas a unha distribución binomial de parámetros $m = 10$ e $p = 0.7$:



Función de masa de probabilidade



Función de distribución

Xeométrica

Novamente, repetindo un mesmo experimento de Bernoulli, pero construíndo a variable aleatoria X que compute o número de fracasos ata o primeiro éxito, obtemos unha variable con distribución xeométrica de parámetro p, e denótase por $X \in Xeo(p)$.

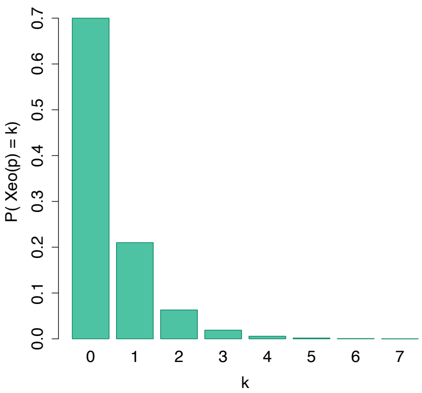
A distribución xeométrica toma como posibles valores $\{0, 1, 2, 3, \ldots\}$ e probabilidades:

$$\mathbb{P}\left(X = k\right) = (1-p)^k p, \text{ con } k = 0, 1, 2, 3, \ldots$$

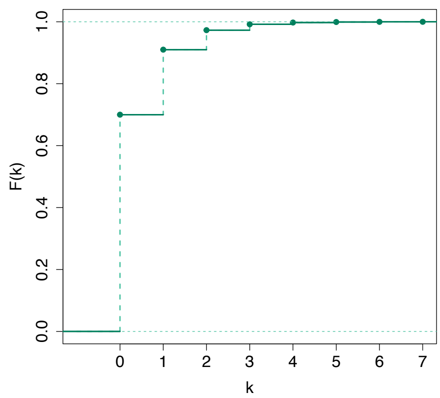
A súa media e a súa varianza veñen dadas por:

$$\mu_X = \frac{1-p}{p} \quad \text{e} \quad \sigma_X^2 = \frac{1-p}{p^2},$$

e representamos a función de masa de probabilidade e a función de distribución asociadas a unha distribución xeométrica de parámetro $p = 0.7$:



Función de masa de probabilidade



Función de distribución

Binomial negativa

A distribución binomial negativa pode verse como unha extensión da xeométrica. Temos de novo un experimento Bernoulli que repetimos nas mesmas condicións, e podemos definir a variable aleatoria X que conte o número de fracasos ata o éxito m-ésimo. Esta variable segue unha distribución binomial negativa de parámetros m e p, e denótase por $X \in BN(m, p)$.

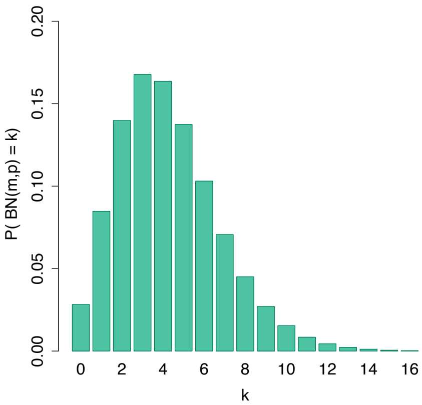
Os posibles valores da distribución binomial negativa son $\{0, 1, 2, 3, \ldots\}$ con probabilidades:

$$\mathbb{P}\left(X = k\right) = \binom{m+k-1}{k} (1-p)^k p^m, \text{ con } k = 0, 1, 2, 3, \ldots$$

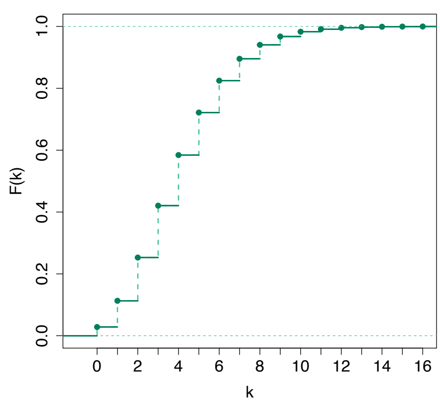
A súa media e varianza veñen dadas por:

$$\mu_X = \frac{m(1-p)}{p} \quad \text{e} \quad \sigma_X^2 = \frac{m(1-p)}{p^2}.$$

A modo de exemplo, representamos a función de masa e a función de distribución asociadas a unha distribución binomial negativa de parámetros $m = 10$ e $p = 0.7$:



Función de masa de probabilidade



Función de distribución

Poisson

Un proceso de Poisson é un experimento aleatorio que consiste en computar o número de sucesos nun certo período de tempo, cumprindo certas hipóteses:

- os sucesos teñen que ocorrer de xeito independente;
- o número medio de eventos por unidade de tempo é constante, e denótase habitualmente por λ.

Nestas condicións, defínese a variable aleatoria $X =$"número de sucesos nun intervalo de tempo", que segue unha distribución de Poisson de parámetro λ, e denótase por $X \in Pois(\lambda)$.

Os posibles valores desta distribución son $\{0, 1, 2, 3, \ldots\}$ con probabilidades dadas por:

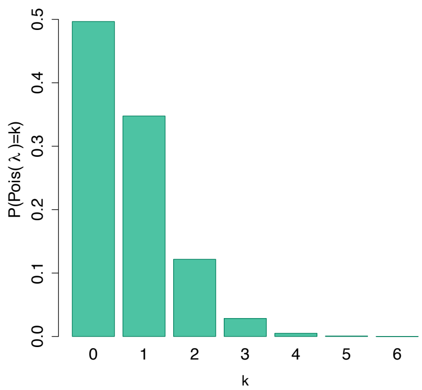
$$\mathbb{P}\left(X = k\right) = \frac{e^{-\lambda}\lambda^k}{k!}, \text{ con } k = 0, 1, 2, 3, \ldots$$

A súa media e varianza coinciden co parámetro que caracteriza a distribución:

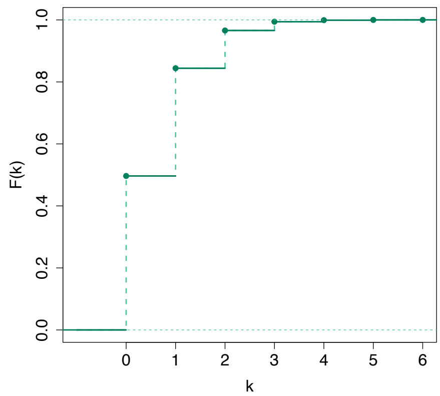
$$\mu_X = \lambda \quad \text{e} \quad \sigma_X^2 = \lambda.$$

Este feito de que media e varianza sexan iguais é especialmente característico da distribución de Poisson. Se tivésemos unha varianza maior falariamos de sobredispersión, e cunha menor de infradispersión.

As funcións características asociadas a unha distribución de parámetro $\lambda = 0.7$ son:



Función de masa de probabilidade



Función de distribución

Como traballamos con variables aleatorias discretas en R ?

Cálculo de probabilidades

`dbinom(k, size = m, prob = p)`	Función de masa de probabilidade dunha $Bi(m, p)$ no punto k.
`pbinom(k, size = m, prob = p)`	Función de distribución dunha $Bi(m, p)$ no punto k.
`dnbinom(k, size = m, prob = p)`	Función de masa de probabilidade dunha $BN(m, p)$ no punto k.
`pnbinom(k, size = m, prob = p)`	Función de distribución dunha $BN(m, p)$ no punto k.
`dpois(k, lambda =` λ`)`	Función de masa de probabilidade dunha $Pois(\lambda)$ no punto k.
`ppois(k, lambda =` λ`)`	Función de distribución dunha $Pois(\lambda)$ no punto k.

Representacións gráficas

`install.packages("LearningStats")`	Descargar e instalar a libraría que precisamos.
`library(LearningStats)`	Cargar na sesión a libraría que precisamos.
`plotBinom(m, p, type = "b")`	Representacións da función de masa e distribución dunha $Bi(m, p)$.
`plotNegBinom(m, p, type = "b")`	Representacións da función de masa e distribución dunha $BN(m, p)$.
`plotPois(lambda, type = "b")`	Representacións da función de masa e distribución dunha $Pois(\lambda)$.

Táboa resumo das distribucións notables discretas

	Soporte	Función de masa	Media	Varianza
Discreta xenérica	$\{x_1, x_2, \dots, x_k\}$	$\{p_1, p_2, \dots, p_k\}$	$\sum_{i=1}^{k} x_i p_i$	$\sum_{i=1}^{k} (x_i - \mu_X)^2 p_i$
Uniforme discreta	$\{x_1, x_2, \dots, x_k\}$	$\mathbb{P}(X = x_k) = \frac{1}{k}$	$\frac{1}{k}\sum_{i=1}^{k} x_i$	$\frac{1}{k}\sum_{i=1}^{k} (x_i - \mu_X)^2$
Bernoulli Ber(p)	$\{0, 1\}$	$\{1-p, p\}$	p	$p(1-p)$
Binomial Bi(m,p)	$\{0, 1, 2, \dots, m\}$	$\mathbb{P}(X = k) = \binom{m}{k} p^k (1-p)^{m-k}$	mp	$mp(1-p)$
Xeométrica Xeo(p)	$\{0, 1, 2, \dots\}$	$\mathbb{P}(X = k) = (1-p)^k p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Binom. Neg. BN(m,p)	$\{0, 1, 2, \dots\}$	$\mathbb{P}(X = k) = \binom{m+k-1}{k} (1-p)^k p^m$	$\frac{m(1-p)}{p}$	$\frac{m(1-p)}{p^2}$
Poisson Pois(λ)	$\{0, 1, 2, \dots\}$	$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$	λ	λ

https://dx.doi.org/10.15304/9788410142077

XUNTA DE GALICIA

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

RAGC/ REAL ACADEMIA GALEGA DE CIENCIAS