Simulação com R

Liane Werner

Simulação do conceito frequentista de probabilidade

- Suponha que o experimento foi repetido n vezes, sempre sob as mesmas condições,
- e que o evento A ocorreu m vezes entre essas n realizações do experimento.
- Então a fração m/n é uma boa aproximação para a probabilidade de A.
- Se o número n de repetições for bastante grande. Simbolicamente, P $(A) \cong n/m$.

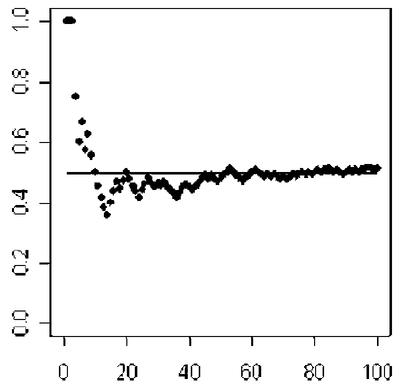
Exemplo: Simulando 100 lançamentos de uma moeda

- No R, foram simulados 100 lançamentos de uma moeda honesta, isto é, onde as chances de cara e de coroa são iguais.
- Depois de cada lançamento, foi observado o número acumulado de caras obtidas até o momento e foi calculada a proporção de caras correspondente.
- A seguir são apresentados os valores correspondentes ao número acumulado de caras ao longo do processo.

1	2	3	3	3	4	4	5	5	5
5	5	5	5	6	7	8	8	9	10
10	10	10	10	11	12	13	13	13	14
14	15	15	15	15	15	16	17	18	18
18	19	20	21	22	22	23	23	23	24
25	26	27	27	27	27	27	28	29	30
31	31	31	31	32	32	33	33	33	34
34	35	36	36	37	38	38	39	39	40
41	41	42	43	43	43	44	44	44	45
46	46	47	47	48	49	50	50	50	51

Exemplo: Simulando 100 lançamentos de uma moeda

- Por exemplo, para a jogada de número 29 o número acumulado de caras é 13 e a fração de caras é 13/29.
- O gráfico mostra a evolução dessa fração a medida que foram feitos os 100 lançamentos da moeda.



Exemplo: Simulando 100 lançamentos de uma moeda

- Observe que no começo há uma grande variabilidade do valor da probabilidade estimada m/n,
- mas ele tende a se estabilizar em 0,5 quando o número n de tentativas vai aumentando.
- a medida que o número n de realizações do experimento aumenta, a probabilidade empírica de um dado evento tende a se estabilizar em uma constante.
- O "ponto de estabilidade" (0,5) corresponde ao valor que seria obtido para a probabilidade de cara, ao usarmos o conceito clássico de probabilidade.

Os comandos no R para a elaboração do gráfico

```
>x=1:100; y=cumsum(sample(0:1,100,rep=T))
>plot(x,y/1:100, ylim=c(0,1), xlim=c(0,100), pch=16)
>segments(1,0.5,100,0.5)
```

Distribuição de probabilidades

- p probabillity – Gera a probabilidade de um valor de x;
- q quantile – Gera o valor x de uma dada probabilidade acumulada;
- d density – Gera o valor da função densidade de um valor x da variável.
- **r randon – Gera n valores do modelo probabilístico em questão.**

Geração de variáveis

- rbinom(n, size, prob) binomial
- rpois(n, lambda) Poisson
- rhyper(nn, m, n, k) hipergeometrica

- runif(n, min=0, max=1) uniforme
- rexp(n, rate=1) exponencial
- rnorm(n, mean=0, sd=1) (normal)

Geração de variáveis

- rt(n, df) 'Student' (*t*)
- rchisq(n, df) Qui-quadrada
- rgamma(n, shape, scale=1) gamma
- rbeta(n, shape1, shape2) beta
- rlnorm(n, meanlog=0, sdlog=1) lognormal
- rweibull(n, shape, scale=1) Weibull

Exercício

- Gerar 1000 valores de uma distribuição normal com média 5 e desvio-padrão 3 e armazenar
- Gerar 1000 valores de uma distribuição exponencial com média 7 e armazenar
- Gerar 1000 valores de uma distribuição weibull com parâmetro de forma 2 e de escala 10 e armazenar
- Plotar o histograma de cada variável gerada
- Somar as distribuições
- Construir o histograma da soma
- Testar se tem distribuição: normal, exponencial, weibull.

- Teste Kolmogorov-Smirnov
  - ks.test (x, "pdistribuição", mean, sd)
  - ks.test (long, "pnorm", mean = mean(long), sd = sqrt(var(long)))

- Distribuição Weibull

$$f(t) = \frac{\gamma}{\theta}\left(\frac{t}{\theta}\right)^{\gamma-1} e^{-\frac{t^{\gamma}}{\theta}}$$

onde γ = parâmetro de forma e θ = parâmetro de escala

$$\text{média} = \theta.\Gamma\left(1+\frac{1}{\gamma}\right) \quad \text{var}= \theta^2.\left[\Gamma\left(1+\frac{2}{\gamma}\right) - \Gamma\left(1+\frac{1}{\gamma}\right)\right]$$