Modelli statistici: sessione 3

Francesco Lagona
Università Roma Tre

1 Variabili aleatorie continue

La funzione di densità di probabilità di una variabile aleatoria (v.a.) continua X è una funzione $f(x) \geq 0$ tale che

$$\int_{-\infty}^{+\infty} f(x)d(x) = 1.$$

La densità di probabilità di X può essere usata per calcolare la probabilità che X assuma valori in un intervallo (a, b):

$$P(a < X < b) = \int_a^b f(x)dx.$$

In particolare, può essere usata per calcolare la funzione di ripartizione di X:

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(u)du.$$

Si osservi che

$$P(a < X < b) = F(b) - F(a)$$

Invertendo la funzione di ripartizione, otteniamo i quantili di X: per ogni $0 \leq p \leq 1$, il quantile di ordine p di X è dato da

$$q_p = F^{-1}(p).$$

Molte v.a. continue che si incontrano nelle applicazioni sono definite in R.

1.1 Variabile aleatoria normale

La variabile aleatoria normale è una delle v.a. più importanti. La sua funzione di densità dipende da due parametri, μ e σ^2, che ne rappresentano rispettivamente il valore atteso e la varianza, e si indica spesso con $N(\mu, \sigma^2)$. In R, il valore della densità normale in x si ottiene con il comando `dnorm`. Ad esempio,

```
> dnorm(0,mean=1,sd=2)
```

```
[1] 0.1760327
```

restituisce il valore in $x = 0$ della densità normale $N(1, 4)$. Il comando

```
> plot(dnorm,from=-3,to=3)
```

disegna invece il grafico di una normale $N(0, 1)$. In R è molto semplice

- calcolare il valore della funzione di ripartizione di una normale

  ```
  > pnorm(0,mean = 1, sd = 2)
  ```

  ```
  [1] 0.3085375
  ```

- calcolare la probabilità di un intervallo sotto una normale

  ```
  > pnorm(2,1,2)-pnorm(1,1,2)
  ```

  ```
  [1] 0.1914625
  ```

- calcolare il quantile di una normale

  ```
  > qnorm(0.975,0,1)
  ```

  ```
  [1] 1.959964
  ```

E' anche possibile estrarre casualmente un campione di osservazioni indipendenti da una popolazione normalmente distribuita:

```
> x <- rnorm(10,1,2)
```

L'istogramma di una campione numeroso è ben approssimato dalla funzione di densità che lo ha generato:

```
> x <- rnorm(10000,0,1)
> hist(x,freq=F)
> plot(dnorm,-3,3,add=T)
```

Se estraiamo un campione di n osservazioni indipendenti da una normale $N(\mu, \sigma^2)$, la v.a. "media campionaria" si distribuisce secondo una normale $N(\mu, \sigma^2/n)$.

```
> m <- rep(0,1000)
> for(i in 1:1000){
+   x <- rnorm(10)
+   m[i] <- mean(x)
+ }
> hist(m,freq=F)
> xx <- seq(-3,3,by=0.01)
> lines(xx,dnorm(xx,0,1/sqrt(10)))
```

2 Variabile aleatoria Chi quadrato

La v.a. Chi Quadrato assume solo valori positivi. La sua densità dipende da un parametro chiamato "gradi di libertà" e si indica con χ^2_n, dove n indica i gradi di libertà. Il seguente comando restituisce il grafico di questa densità per $n = 1, 2, 10$:

```
> xx <- seq(0,20,by=0.01)
> plot(xx,dchisq(xx,1),type="l",ylim=c(0,0.6))
> lines(xx,dchisq(xx,2),col=2)
> lines(xx,dchisq(xx,10),col=4)
```

I comandi `pchisq`, `qchisq` e `rchisq` restituiscono rispettivamente la funzione di ripartizione, il quantile e una valore casuale di una v.a. Chi Quadrato.

Se estraiamo un campione di n osservazioni indipendenti da una normale $N(\mu, \sigma^2)$ e consideriamo la varianza campionaria corretta

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

la v.a.

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

. Possiamo verificare questo risultato tramite simulazione:

```
> v <- rep(0,1000)
> for(i in 1:1000){
+   x <- rnorm(10)
+   v[i] <- 9*var(x)/1
+ }
> hist(v,freq=F)
> xx <- seq(0,30,by=0.01)
> lines(xx,dchisq(xx,9))
```

3 Variabile aleatoria t di Student

La densità della v.a. t di Student, t_n è simmetrica rispetto a 0 e la sua forma dipende dal parametro n che prende il nome di "gradi di libertà". Il seguente comando restituisce il grafico di questa densità per $n = 1, 2, 10$:

```
> xx <- seq(-3,3,by=0.01)
> plot(xx,dt(xx,1),type="l",ylim=c(0,0.5))
> lines(xx,dt(xx,2),col=2)
> lines(xx,dt(xx,10),col=4)
```

I comandi `pt`, `qt` e `rt` restituiscono rispettivamente la funzione di ripartizione, il quantile e una valore casuale di una v.a. t di Student.

Se due variabili indipendenti si distribuiscono rispettivamente secondo una normale standardizzata e un chi quadrato con $n-1$ gradi di libertà (gdl),

$$X \sim N(0,1) \quad Y \sim \chi^2_{n-1}$$

allora il rapporto

$$\frac{X}{\sqrt{Y/(n-1)}} \sim t_{n-1}.$$

Di conseguenza, se estraiamo un campione di n osservazioni indipendenti da una normale $N(\mu, \sigma^2)$, la v.a.

$$\frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)\hat{\sigma}^2}{\sigma^2}}{n-1}}} = \frac{\bar{x}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$$

4 Variabile aleatoria F di Fisher

Se abbiamo a che fare con due variabili indipendenti

$$X \sim \chi^2_n \quad Y \sim \chi^2_m$$

allora il rapporto

$$\frac{X/n}{Y/m} \sim F_{n,m}$$

dove $F_{n,m}$ è la densità di una variabile aleatoria di Fisher con parametri n ed m. I comandi `df`, `pf`, `qf` e `rf` restituiscono respettivamente la densità, la funzione di ripartizione, il quantile e il valore casuale di una v.a. F.