Lezione n. 3

3.1 Grafici e distribuzioni

Esempio 3.1 *Distribuzione Gamma*
Supponiamo che $X \sim \text{Gamma}(\alpha, \beta)$. In questo caso abbiamo

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x) \;\; x > 0, \;\; \alpha, \lambda > 0 \tag{3.1}$$

Vedremo di rappresentare graficamente:

(i) la densità precedente, per diverse scelte dei parametri;

(iii) la distribuzione di somme di variabili gamma, $S_n = \sum_{i=1}^n Y_i$, con $Y_i, i = 1, \cdots, n$ indipendenti e distribuite secondo una legge $Gamma(\lambda, \alpha_i)$. In questo caso considereremo anche la distribuzione empirica, rappresentandola tramite l'istogramma costruito su una simulazione dalla distribuzione di S_n. In particolare si analizzerà il caso in cui le variabili Y_i sono di tipo chi quadrato.

(i) Prima di rappresentare le densità di alcune particolari Gamma, ricordiamo che per $\alpha = 1$ la densità 3.1 diventa ($\Gamma(1) = 1$)

$$f(x; \alpha = 1, \lambda) = \lambda \exp(-\lambda x) \;\; x > 0, \;\; \lambda > 0,$$

che è la densità di un'esponenziale negativa di parametro λ. Per la scelta di parametri $\lambda = \nu/2, \alpha = 1/2$ la 3.1 assume invece la forma seguente:

$$f(x; \alpha = \frac{\nu}{2}, \lambda = \frac{1}{2}) = \frac{(\frac{1}{2})^{\nu/2}}{\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \;\; x > 0, \nu \in \mathcal{N}$$

che è la densità di un Chi quadrato con ν gradi di libertà.

Per disegnare con R le densità Gamma, ivi comprese Esponenziale negativa e Chi quadrato, ricordiamo che il programma ha incorporate le funzioni che restituiscono le densità di molte distribuzioni note; tra queste anche la Gamma.
Per quanto riguarda la famiglia di densità Gamma, esistono due parametrizzazioni equivalenti, che differiscono per il secondo parametro. Una è basata sul cosiddetto tasso (rate) λ come nella parametrizzazione usata in 3.1, l'altra è basata sul cosiddetto parametro di scala (scale) $\gamma = 1/\lambda$. R ha la funzione

```
dgamma(x, shape, rate, scale=1/rate)
```

che calcola nel punto x la densità di una Gamma di parametri assegnati. La presenza di tre parametri (oltre all'argomento `x`) si deve al fatto che R pemette di utilizzare l'una o l'altra delle parametrizzazioni, potendosi specificare il parametro `rate=`λ oppure il parametro `scale=`$1/\lambda$. Il primo parametro (α nella notazione usata finora) corrisponde a `shape`.

Nello specificare gli argomenti di una funzione, in R possiamo omettere i nomi degli argomenti, nel qual caso conta l'ordine con cui essi sono introdotti. Ad esempio è equivalente scrivere `gamma(shape=1, rate=1/2)` oppure `gamma(1,1/2)`, mentre se intendiamo ricorrere alla parametrizzazione con il parametro di scala pari a 1/2 dovremo necessariamente scrivere `gamma(1, sacale=1/2)` (oppure, naturalmente, `gamma(1,2)`).

Per disegnare le densità utilizziamo la funzione `curve` che ha la seguente sintassi:

```
curve(expr,from,to,n=101,add=FALSE,type="l",ylab=NULL,...).
```

Essa permette di disegnare grafici di funzioni, siano esse disponibili in R o costruite direttamente dall'utente. La funzione crea automaticamente una successione di punti in cui valutare la funzione o espressione `expr` da disegnare. `from` e `to` indicano gli estremi dei punti su cui la funzione viene valutata, mentre `n` il cui default è 101, premette di indicare il numero complessivo di punti su cui valutare la funzione.

Applicando `curve` alla nostra funzione di densità `dgamma`, la sintassi è:

```
curve(dgamma(x, 1,10),0.01,25,n=1000)
```

Avremo il grafico di una densità Esponenziale negativa di media 1/10, rappresentato con una linea continua ("solida"). Avremmo lo stesso tipo di grafico creando una sequenza di valori su cui valutare la funzione e disegnando la curva per punti con i comandi seguenti:

```
y <- seq(0.01,25,0.1)
plot(y, dgamma(y,1,10), type='l')
```

(`type='l'` indica che i punti sul grafico vanno congiunti con una linea smussata)

Per *sovrapporre* i grafici di una densità Gamma(12,10) ed una di tipo Chi quadrato di parametro 4 scriviamo invece

```
curve(dgamma(x, 12,10),0.01,25, lty=2)
curve(dgamma(x, 2,1/2),n=1000, lty=3, add=T)
```

`lty` indica il parametro grafico line type, l'intero che gli si assegna determina il tipo di tratteggio. Specificare `add=T` nel comando `curve` implica che la curva verrà disegnata sul grafico precedente, senza creare un nuovo grafico.

La sovrapposizione con la densità esponenziale negativa produce risultati esteticamente poco soddisfacenti. Meglio scrivere i seguenti comandi alternativi:

```
plot(y,dgamma(y,12,10),type='l')
lines(y,dgamma(y,2,1/2),lty=2)
lines(y,dgamma(y,1,10),lty=3)
```

`points` è un comando di secondo livello, ossia ha effetto sul grafico corrente e soltanto se c'é una finestra grafica aperta. Corrisponde sostanzialmente al comando `plot` con il parametro `type='l'`, con la differenza che `plot` é un comando di primo livello, e quindi quando viene invocato determina la creazione di un nuovo grafico o di una finestra grafica se non ve ne é una aperta.

(ii) Per determinare la distribuzione della somma di v.a. Gamma indipendenti con lo stesso tasso λ, utilizziamo la funzione generatrice dei momenti di una Gamma di parametri λ, α, $G_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$. Sfruttando l'indipendenza tra le Y_i, la funzione generatrice dei momenti di $\sum_{i=1}^{n} Y_i$ con Y_i *ind.* $\sim$ Gamma(λ, α_i) è pari a:

$$G_Y(t) = \left\{\frac{\lambda}{\lambda - t}\right\}^{\sum_{i=1}^{n} \alpha_i}$$

che è anche la f.g.m. di una Gamma$(n, \sum_{i=1}^{n} \alpha_i)$.

Questo risultato può essere particolarizzato per mostrare che la somma di variabili chi quadrato indipendenti è distribuita ancora come un Chi quadrato con parametro pari alla somma dei gradi di libertà delle variabili sommate.

Osserviamo graficamente, tramite la rappresentazione della distribuzione empirica (istogramma) di un campione di 5000 unità, che il quadrato di una variablie aleatoria Normale standard ha distribuzione Chi(1). A questo scopo estraiamo un campione di 5000 unità da una legge N(0,1) e calcoliamo il quadrato dei valori ottenuti. Per verificare con lo stesso strumento anche il risultato dimostrato precedentemente relativo alla somma di variabili Chi quadrato indipendenti, creiamo inoltre una funzione che somma k vettori (indipendenti perché estratti indipendentemente) dei quadrati dei valori estratti da una N(0,1). Otterremo per $k = 1$ un campione da un Chi(1), per $k = 2$ un campione da un Chi(2) e così via. Per verificare questo asserto, sovrapponiamo inoltre all'istogramma la funzione di densità appropriata (ossia quella di un Chi quadrato con k gradi di libertà). Si noti che per disegnare un istogramma confrontabile con una funzione di densità la somma delle aree dei rettangoli dell'istogramma deve essere pari a uno, quindi non deve essere uguale al numero totale di unità rappresentate. A questo scopo occorre specificare l'opzione `prob=T`. In sostanza l'opzione `prob=T` implica che nell'istogramma le aree rappresentano le frequenze campionarie *relative*, menre l'opzione `prob=F` (che é il default) implica che le frequenze campionarie rappresentate sono quelle assolute.

```
chiq_function (k=2,n=5000)
{
ck_rep(0,n)
for (i in 1:k){ck_ck+(rnorm(n))^2}
ak_sort(ck)
hist(ak,prob=T,nclass=40)
lines(ak,dchisq(ak,k),col=3)
return(ck)
}
```

Scrivendo quindi ad esempio uno dei (o ambedue i) comandi

```
ak<-chiq(1)
bk<-chiq(16)
```

otterremo sulla finestra grafica l'istogramma di un Chi(1) (un Chi(16)), mentre il campione di 5000 unità (che é il valore di default nella nostra funzione) estratto da una legge Chi(1) (Chi(16)) é salvato in `ak` (`bk`).

ESEMPIO 3.2 *Distribuzione t di Student*
Con lo strumento della simulazione, ossia utilizzando campioni di numerosità sufficientemente elevata e rappresentandone graficamente la distribuzione, ad esempio tramite un istogramma, è possibile investigare il tipo di legge indotto da determinate trasformazioni il cui studio analitico potrebbe essere complicato. Nell'esempio precedente abbiamo già utilizzato trasformazioni di variabili aleatorie Normali standard indipendenti, nella fattispecie $\sum_{i=1^k} X_i^2$. Un procedimento del tutto analogo può essere seguito per verificare che la trasformata $T = \frac{Z}{\sqrt{Y/\nu}}$, con $Z \sim N(0,1)$ e $Y \sim \chi^2(\nu)$ indipendenti, segue una legge di Student con ν gradi di libertà. Estraiamo indipendentemente un campione da una Normale standard (comando `rnorm`) ed un campione da un Chi quadrato con 16 g.d.l. (funzione `chiq()` costruita precedentemente e già utilizzata per generare un campione di 5000 unità da un Chi(16), salvato in `ak`, oppure comando `rchisq()`; ovviamente i campioni devono essere della stessa numerosità. Per ogni $i = 1, \ldots 5000$, cioé per ogni elemento dei vettori, costruiamo il rapporto $Z_i/\sqrt{Y_i/\nu}$: ricordando che R opera sui vettori applicando l'aritmetica *elemento per elemento*, scriviamo

```
ti <- rnorm(5000)/(sqrt(ak/16))
ti <- sort(ti)
hist(ti, prob=T, nclass=80)
lines(ti,dnorm(ti),col=4)
lines(ti,dt(ti,16),col=3)
```

Dal confronto con la densità normale standard si verifica facilmente che la distribuzione t ha le code più pesanti della normale. Questo si verifica anche confrontando numericamente le probabilità relative alle code delle due distribuzioni, ad esempio

```
> pnorm(-3)
[1] 0.001349898
> pt(-3,16)
[1] 0.00423975
```

per ottenere un'*approssimazione* della precedente probabilità sotto la distribuzione t sfruttiamo il campione: `sum(ti<=-3)/length(ti)`. Confrontare questo risultato con quello teorico (`pt(-3,16)`) anche al variare della numerosità del campione generato.

Esercizio 3.1. *Asimmetria delle densità Gamma*
Fissato $\lambda = 1$, disegnare le densità Gamma per $\alpha \in \{1, 2, 3, 13, 20\}$ rispettivamente. Utilizzando il comando `curve` si può scrivere

```
curve(dgamma(x,1,1),0.01,30,ylim=c(0,0.5),n=1000)
curve(dgamma(x,2,1),n=1000,add=T,col=2)
curve(dgamma(x,3,1),n=1000,add=T,col=3)
curve(dgamma(x,13,1),n=1000,add=T,col=4)
curve(dgamma(x,20,1),n=1000,add=T,col=5)
title("densita' Gamma")
```

In questo caso, il parametro `col` permette di specificare il *colore* della curva.

Si noti che le densità Gamma sono asimmetriche, e che tale caratteristica tende ad attenuarsi quando α cresce. Guardiamo più "da vicino" l'ultima delle densità disegnate, scrivendo `curve(dgamma(x,20,1),n=1000,0.01,40)`.
Anche scrivendo `curve(dgamma(x,50,1),n=1000,0.01,80)`, possiamo facilmente constatare che l'asimmetria della distribuzione si mantiene anche quando α è piuttosto grande.