Un tipico esame consiste di tre domande del tipo sottostante:

1) Una situazione standard da analizzare:

- un modello genetico suggerisce che la composizione genetica di una data popolazione in geni AA, Aa e aa sia di tipo binomiale Bin(2,p) (p parametro da stimare). In un campione casuale di 100 individui si riscontrano 40 di tipo AA, 20, di tipo Aa e 40 di tipo aa. A quale conclusione portano questi dati?

2) Sia $X_1,\ldots X_n$ iid con distribuzione Exp(λ). "Tutti" sanno che lo stimatore ML di $1/\lambda$ e' la media $\bar{X}$ delle osservazioni... Si vuole pero' stimare $P(X>1)=e^{-\lambda}$ e si pensa di usare $\exp(-1/\bar{X})$ come stimatore. Qual e' la distribuzione asintotica di questo stimatore?
Un'altra possibilita' e' di calcolare la proporzione $\hat{p}$ di osservazioni con valore maggiore di 1. Questo e' anche uno stimatore di $P(X>1)=e^{-\lambda}$. Qual e' la varianza asintotica di questo stimatore? Quale dei due ha la varianza minore?

3) Se si fanno due misure X e Y indipendenti da una normale $N(\mu,\sigma^2)$, la stima $\hat{\mu}$ della media μ e' (X+Y)/2 e quella $\hat{\sigma}$ di σ e' $\sqrt{(X-Y)^2/2}$. Quale grado di confidenza e' da attribuire all'intervallo $\hat{\mu}\pm 1.96\hat{\sigma}/\sqrt{2}$? (si puo' rispondere anche in modo simbolico, se la tabella apposita non e' disponibile).

Cose forse utili:

```
> qchisq(0.95,1)
[1] 3.841459
> qt(0.975,98)
[1] 1.984467
> pt(1.96,1)
[1] 0.8498286
```

4) Due situazioni standard da analizzare:

- in un negozio, si studiano i primi 100 clienti che entrano e li si classifica secondo sesso (M/F) e eta' (giovane/meno giovane... G/V). Il risultato e' MG = 35, MV=35, FG=20, FV=10.Vi e' evidenza statisticamente significativa per una differenza tra maschi e femmine per quanto riguarda l'eta'?

- in realta', nello studio sopra, si sa l'eta' precisa delle persone osservate: la media per maschi e' 35 anni con sd campionaria 10 e, per femmine, media 25 e sd campionaria anche 10...Questi dati portano a una conclusione differente da quella ottenuta sopra?

5) Sia (A,B,C) una distribuzione multinomiale basata su n ripetizioni di un esperimento con tre possibili esiti con probabilita' (p,2p,1-3p) per p in [0,1/3]. Trovare lo stimatore ML di p.

6) Due situazioni standard da analizzare:

- in due popolazioni diverse, vengono estratti campioni casuali di grandezza 50 e sui soggetti estratti viene determinata la presenza o meno di un certo gene... I risultati sono: pop. A (si=32, no=18) e pop. B (si = 22, no=28). Vi e' evidenza statisticamente significativa per una differenza tra le popolazioni?

- per controllare se un dato strumento e' ben calibrato, si misura tre volte un oggetto standard di valore noto 1. Le misurazioni hanno media 1.1 e deviazione standard stimata 0.3. Assumendo distribuzione normale per la variazioni dello strumento, vi e' evidenza statisticamente significativa per un bias sistematico delle misure?

7) Sia $X_1,\ldots X_n$ iid con distribuzione Poisson(λ). "Tutti" sanno che lo stimatore ML di λ e' la media $\bar{X}$ delle osservazioni... Si vuole pero' stimare $\sqrt{\lambda}$ e si pensa di usare $\sqrt{\bar{X}}$ come stimatore. Qual e' la distribuzione asintotica di questo stimatore?

8) Dimostrare che le tre seguenti funzioni di ripartizione possono essere ottenute tramite opportune trasformazioni di una variabile con distribuzione Exp(1) e completare la loro definizione su tutto l'asse reale:

- $F(x) = 1 - e^{-\lambda x}$

- $F(x) = 1 - e^{-\lambda x^k}$

- $F(x) = e^{-e^{-\lambda x}}$

9) Si dice che una distribuzione t con k gradi di liberta' e' piu' "diffusa" o "piatta" di una distribuzione normale standard. Questo vuol dire, per ogni x>0, che P(N>x) < P(T>x), dove N e T sono due variabili aleatorie con le rispettive distribuzioni normale standard e t con k gradi di liberta'...
Dimostrare che questo e' vero, per esempio usando la disuguaglianza di Jensen e la definizione di una distribuzione t come N/sqrt(C/k) (cosa e come devono essere N, C e k in questa definizione?)... Iniziare dall'espressione P(T≤x) = E(F(x sqrt(C/k))), dove F rappresenta la funzione di ripartizione della normale standard, dopo aver spiegato da dove viene quest'espressione...

10) Sia X_1,...,X_n) iid N(\mu,\sigma^2) e s^2 la varianza campionaria.
Trovare la distribuzione asintotica (per n grande) di s, la radice quadrata di s^2

11) Sia X_1,...,X_n iid Bernoulli(p) e l'ipotesi nulla H_0: p=1/2 vs HA:p≠1/2. Trovare il test LR e esprimerlo nel modo piu' semplice possibile...

12) Sia $X_1,...X_n$ iid con funzione di ripartizione $F(x)=1-\exp(-\theta x)$, $0<\theta$.

a) Trovare lo stimatore ML di θ e la sua distribuzione asintotica.

b) Si puo' ottenere una v.a. distribuita come sopra tramite una semplice trasformazione di una v.a. con distribuzione uniforme?

c) Supponendo di volere anche stimare $\alpha = P(X \leq 1)=1-\exp(-\theta)$, come si calcolerebbe lo standard error stimato (asintotico) della stima?

13) Sia X e Y due v.a. iid N(0,1). Sia A=X+Y e B= X+2Y.

a) Determinare se (A,B) ha una distribuzione bivariata normale e, semmai, determinarne i parametri.

b) Calcolare E(B | A=s).

14) Sia X distribuito come Bin(n,p). Qual e' la forma dell'intervallo di confidenza per p, usando l'approssimazione normale per la binomiale che viene usata di solito (intervallo di Wald)? Agresti e Coull propongono, dopo attenti studi, una formula alternativa che, secondo

loro, approssima meglio il grado di confidenza. Cosa puo' volere dire questo? La loro formula per un intervallo con grado di confidenza 95% consiste in calcolare p*=(X+2)/n+4 e poi calcolare l'intervallo p*±zs/sqrt(n+4) con s=sqrt(p*(1-p*)) e z=1.96. Dimostrare che per grande valore di n, le due formule saranno equivalenti...

15) Sia $X_1,\ldots X_n$ iid con funzione di ripartizione $F(x)=1-\exp(-\theta x^2)$, $0<\theta$.

a) Trovare lo stimatore ML di θ e la sua distribuzione asintotica.

b) Si puo' ottenere una v.a. distribuita come sopra tramite una semplice trasformazione di una v.a. con distribuzione esponenziale?

16) Sia X e Y due v.a. iid N(0,1). Sia A=X e B= X+Y.

a) Dimostrare che (A,B) ha una distribuzione bivariata normale e determinarne i parametri e la densita'.

b) Calcolare la probabilita' che $\{A>0 \cap B>0\}$ (hint: ritornare a (X,Y) e pensare in coordinate polari...)

17) Si lancia una moneta 10 volte e si denota il numero di Testa ottenuti X. Si vuole testare l'ipotesi nulla che p=1/2. Si rifiuta l'ipotesi nulla se X = 9 o 10. Calcolare il livello di questo test (α) e la potenza del test per p=1/3 e 2/3.

18) Sia $X_1,\ldots X_n$ iid con la distribuzione discreta che assegna probabilita' $(\theta/2,1-\theta,\theta/2)$, $0\leq\theta\leq1$ ai tre valori (1,2,3). Trovare lo stimatore ML di θ e la sua distribuzione asintotica.

19) Sia X una v.a. con distribuzione $\mathrm{Exp}(\theta)$, $\theta>0$, cioe' con densita' $f(x)=\theta\exp(-\theta x)$, $x\geq 0$.

a) Dimostrare che θX ha una distribuzione indipendente da θ.

b) Usare questo risultato per trovare un intervallo di confidenza di grado $1-\alpha$, $0<\alpha<1$, per θ.

c) Quale scelta di limiti minimizza la lunghezza dell'intervallo?

d) Come si estende i risultati sopra a un campione $X_1, \ldots X_n$ iid Exp(θ)?

20) Due studiosi di erpetometria studiano i loro animali favoriti in due localita' diverse, classificandone la colorazione (Chiaro o Scuro). Il primo ha un campione di numerosita' 50, di cui 22 Scuri, e il secondo un campione di numerosita' anche 50, con ben 28 Scuri. E' statisticamente significativa la differenza? Specificare qual e' l'ipotesi sotto test e qual e' la conclusione del confronto.