Richiami sulle distribuzioni campionarie (F. De Santis)

1. **Principali statistiche campionarie**[1]

Statistica	$T(\mathbf{X}_n)$	Funzione
Media campionaria	$\overline{X}_n$	$\frac{1}{n}\sum_{i=1}^n X_i$
Somma campionaria	Y_n	$\sum_{i=1}^n X_i = n\overline{X}_n$
Varianza campionaria	$\widehat{\sigma}_n^2$	$\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
Varianza campionaria corretta	S_n^2	$(n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
Minimo campionario	$X_{(1)}$	$\min\{X_1, \ldots, X_n\}$
Massimo campionario	$X_{(n)}$	$\max\{X_1, \ldots, X_n\}$

2. **Valore atteso e varianza di somma, media e varianza campionarie**

   $X_1, \ldots, X_n$ iid con $\mathbb{E}[X] = \mu$ e $\mathbb{V}[X] = \sigma^2 < \infty$. Per le statistiche campionarie $\overline{X}_n$, $\widehat{\sigma}_n^2$ e S_n^2 si ha:

   - $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X] = \mu$, $\quad \mathbb{V}[\overline{X}_n] = \frac{1}{n}\mathbb{V}[X] = \frac{\sigma^2}{n}$ e quindi $\mathbb{E}[\overline{X}_n^2] = \mathbb{V}[\overline{X}_n] + (\mathbb{E}[\overline{X}_n])^2 = \frac{\sigma^2}{n} + \mu^2$;
   - $\mathbb{E}[\widehat{\sigma}_n^2] = \dfrac{n-1}{n}\mathbb{V}[X] = \dfrac{n-1}{n}\sigma^2$;
   - $\mathbb{E}[S_n^2] = \mathbb{V}[X] = \sigma^2$.

3. **Campionamento da popolazioni Bernoulli e Poisson**

   - $X_i \sim \text{Ber}(p)$ iid $i = 1, \ldots, n \Longrightarrow \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$
   - $X_i \sim \text{Pois}(\lambda)$ iid $i = 1, \ldots, n \Longrightarrow \sum_{i=1}^n X_i \sim \text{Pois}(n\lambda)$

4. **Campionamento da popolazioni normali**

   - $X_i \sim \text{N}(\mu, \sigma^2)$ iid $i = 1, \ldots, n \Longrightarrow$
     - $\sum_{i=1}^n X_i \sim \text{N}(n\mu, n\sigma^2)$
     - $\overline{X}_n \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$
     - $aX + b \sim \text{N}(a\mu + b, b^2\sigma^2)$
   - Proprietà della v.a. Chi quadrato
     - $Z \sim \text{N}(0,1) \Longrightarrow Z^2 \sim \chi_1^2$
     - $Z_1, \ldots, Z_k$ indipendenti $\text{N}(0,1) \Longrightarrow \sum_{i=1}^k Z_i^2 \sim \chi_k^2$
     - $X_i \sim \chi_{p_i}^2$ indipendenti, $p_i > 0$, $i = 1, \ldots, n \Longrightarrow \sum_{i=1}^n X_i \sim \chi_p^2$, con $p = \sum_{i=1}^n p_i$
     - $X \sim \chi_p^2 \Longrightarrow \mathbb{E}[X] = p$, $\mathbb{V}[X] = 2p$

[1] **Importante**: qui le distribuzioni di tutte le v.a. sono distribuzioni campionarie e quindi *condizionate* ai parametri. Omettiamo di indicarlo per brevità.

- $X_i \sim \mathrm{N}(\mu_0, \sigma^2)$ iid (μ_0 noto), $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2 \implies$
  - $S_0^2 \sim \mathrm{Ga}\left(\frac{n}{2}, \frac{n}{2\sigma^2}\right)$ e $\frac{nS_0^2}{\sigma^2} \sim \chi_n^2$
  - $\mathbb{E}[S_0^2] = \sigma^2$, $\mathbb{V}[S_0^2] = \frac{2\sigma^4}{n}$
- $X_i \sim \mathrm{N}(\mu, \sigma^2)$ iid, $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ e $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \implies$
  - $S_n^2 \sim \mathrm{Gamma}\left(\frac{n-1}{2}, \mathtt{rate} = \frac{n-1}{2\sigma^2}\right)$, $\quad \mathbb{E}[S_n^2] = \sigma^2 \quad$ e $\quad \mathbb{V}[S_n^2] = \frac{2\sigma^4}{n-1}$,
  - $\widehat{\sigma}_n^2 \sim \mathrm{Gamma}\left(\frac{n-1}{2}, \mathtt{rate} = \frac{n}{2\sigma^2}\right)$, $\quad \mathbb{E}[\widehat{\sigma}_n^2] = \frac{n-1}{n}\sigma^2 \quad$ e $\quad \mathbb{V}[\widehat{\sigma}_n^2] = \frac{2(n-1)\sigma^4}{n^2}$,
  - $\frac{(n-1)S_n^2}{\sigma^2} = \frac{n\widehat{\sigma}_n^2}{\sigma^2} \sim \chi_{n-1}^2$
- t di Student
  - $U \sim \mathrm{N}(0,1)$ e $V \sim \chi_p^2$ indipendenti $\implies T = \frac{U}{\sqrt{\frac{V}{p}}} = \frac{\sqrt{p}U}{\sqrt{V}} \sim t_p$ (t di Student con p gradi di libertà)
  - $X_i \sim \mathrm{N}(\mu, \sigma^2)$ iid $\implies U = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim \mathrm{N}(0,1)$ e $V = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$
  - U e V indipendenti
  - $T = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}$

5. **Campionamento da popolazioni gamma**

- $Z \sim \mathrm{Gamma}(\alpha, \mathtt{rate} = \beta)$ e $a > 0 \quad \implies \quad aZ \sim \mathrm{Gamma}(\alpha, \mathtt{rate} = \beta/a)$.
- $Z \sim \mathrm{Gamma}(\alpha, \mathtt{scale} = \beta)$ e $a > 0 \quad \implies \quad aZ \sim \mathrm{Gamma}(\alpha, \mathtt{scale} = a\beta)$.
- Relazioni

X_i	$Y_n = \sum_{i=1}^n X_i$	$\overline{X}_n$
Gamma $(\alpha, \mathtt{rate} = \beta)$	$\mathrm{Gamma}(n\alpha, \mathtt{rate} = \beta)$	$\mathrm{Gamma}(n\alpha, \mathtt{rate} = n\beta)$
$\mathrm{EN}(\beta) = \mathrm{Gamma}(1, \mathtt{rate} = \beta)$	$\mathrm{Gamma}(n, \mathtt{rate} = \beta)$	$\mathrm{Gamma}(n, \mathtt{rate} = n\beta)$
Gamma $(\alpha, \mathtt{scale} = \beta)$	$\mathrm{Gamma}(n\alpha, \mathtt{scale} = \beta)$	$\mathrm{Gamma}(n\alpha, \mathtt{scale} = \beta/n)$
$\mathrm{Esp}(\beta) = \mathrm{Gamma}(1, \mathtt{scale} = \beta)$	$\mathrm{Gamma}(n, \mathtt{scale} = \beta)$	$\mathrm{Gamma}(n, \mathtt{scale} = \beta/n)$

6. **Distribuzione di $X_{(n)}$ e di $X_{(1)}$**

- $X_1, \ldots, X_n$ n iid, ciascuna con funzione di ripartizione $F_X(\cdot;\theta)$
  Allora per $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$ si ha
  -$F_{X_{(1)}}(x) = 1 - [1 - F_X(x;\theta)]^n$
  - $F_{X_{(n)}}(x) = [F_X(x;\theta)]^n$
  - $f_{X_{(1)}}(x) = n[1 - F_X(x;\theta)]^{n-1} f_X(x,\theta)$
  - $f_{X_{(n)}}(x) = n[F_X(x;\theta)]^{n-1} f_X(x;\theta)$
- $X_1, \ldots, X_n$ iid $\mathrm{Unif}[0,\theta] \implies$
  - $F_X(x) = \frac{x}{\theta} I_{[0,\theta]}(x)$
  - $F_{X_{(n)}}(x) = \left[F_X(x)\right]^n = \left[\frac{x}{\theta}\right]^n I_{[0,\theta]}(x)$
  - $f_{X_{(n)}}(x) = n\left[\frac{x}{\theta}\right]^{n-1} \frac{1}{\theta} I_{[0,\theta]}(x) = n\frac{x^{n-1}}{\theta^n} I_{[0,\theta]}(x)$
  - $\mathbb{E}(X_{(n)}) = \frac{n}{n+1}\theta$
  - $\mathbb{E}(X_{(n)}^2) = \frac{n}{n+2}\theta^2$
  - $\mathbb{V}(X_{(n)}) = \frac{n\theta^2}{(n+1)^2(n+2)}$

7. **Approssimazioni asintotiche delle d. campionarie di $\overline{X}_n$ e Y_n**

   - $X_1, \ldots, X_n$ iid con $\mathbb{E}[X]$ e $\mathbb{V}[X] \Longrightarrow$
     $\overline{X}_n \dot{\sim} \mathrm{N}\left(\mathbb{E}[X], \frac{\mathbb{V}[X]}{n}\right)$ e $Y_n = \sum_{i=1}^n X_i \dot{\sim} \mathrm{N}\left(n\mathbb{E}[X], n\mathbb{V}[X]\right)$

8. **Distribuzione asintotica stimatori di massima verosimiglianza** (modelli regolari)

   - Se d_{mv} e d_{mv}^g sono gli smv rispettivamente di θ e $g(\theta)$ allora [se esiste $g'(\cdot)$]
     - $d_{mv} \dot{\sim} \mathrm{N}(\theta, I_n^{-1}(\theta))$
     - $d_{mv}^g \dot{\sim} \mathrm{N}(g(\theta), [g'(\theta)]^2 I_n^{-1}(\theta))$