Richiami sulle distribuzioni campionarie (F. De Santis)

1. **Principali statistiche campionarie**

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	$\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
Varianza camp. (con μ_0 noto)	S_0^2	$\frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$
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 Y_n, $\overline{X}_n$, $\widehat{\sigma}_n^2$ e S_n^2 si ha:

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

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

   - $X_i \sim \text{Ber}(p)$ iid $i = 1, \ldots, n \Longrightarrow Y_n = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$
     - $\mathbb{E}[Y_n] = np, \quad \mathbb{V}[Y_n] = np(1-p)$
     - $\mathbb{E}[\overline{X}_n] = p, \quad \mathbb{V}[\overline{X}_n] = \frac{p(1-p)}{n}$
   - $X_i \sim \text{Pois}(\lambda)$ iid $i = 1, \ldots, n \Longrightarrow Y_n = \sum_{i=1}^n X_i \sim \text{Pois}(n\lambda)$
     - $\mathbb{E}[Y_n] = n\lambda, \quad \mathbb{V}[Y_n] = n\lambda$
     - $\mathbb{E}[\overline{X}_n] = \lambda, \quad \mathbb{V}[\overline{X}_n] = \frac{\lambda}{n}$

4. **Campionamento da popolazioni normali**

   - $X_1, \ldots, X_n \sim \text{N}(\mu, \sigma^2)$ iid $\Longrightarrow$
     - $\sum_{i=1}^n X_i \sim \text{N}(n\mu, n\sigma^2)$
     - $\overline{X}_n \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$
     - $aX + b \sim \text{N}(a\mu + b, a^2\sigma^2)$

- Proprietà della v.a. chi quadrato
  - $Z \sim \mathrm{N}(0,1) \Longrightarrow Z^2 \sim \chi^2_1$
  - $Z_1, \ldots, Z_k$ indipendenti $\mathrm{N}(0,1) \Longrightarrow \sum_{i=1}^k Z_i^2 \sim \chi^2_k$
  - $X_i \sim \chi^2_{p_i}$ indipendenti, $p_i > 0$, $i = 1, \ldots, n \Longrightarrow \sum_{i=1}^n X_i \sim \chi^2_p$, con $p = \sum_{i=1}^n p_i$
  - $X \sim \chi^2_p$, $p > 0 \Longrightarrow \mathbb{E}[X] = p$, $\mathbb{V}[X] = 2p$
  - $\chi^2_p \equiv \mathrm{Ga}\left(\frac{p}{2}, \texttt{rate} = \frac{1}{2}\right)$

- $X_1, \ldots, X_n \sim \mathrm{N}(\mu_0, \sigma^2)$ iid (μ_0 noto), $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2 \Longrightarrow$
  - $S_0^2 \sim \mathrm{Ga}\left(\frac{n}{2}, \frac{n}{2\sigma^2}\right)$ e $\frac{nS_0^2}{\sigma^2} \sim \chi^2_n$
  - $\mathbb{E}[S_0^2] = \sigma^2 \qquad \mathbb{V}[S_0^2] = \frac{2\sigma^4}{n}$

- $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ iid, $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ e $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \Longrightarrow$
  - $S_n^2 \sim \mathrm{Gamma}\left(\frac{n-1}{2}, \texttt{rate} = \frac{n-1}{2\sigma^2}\right)$, $\mathbb{E}[S_n^2] = \sigma^2$ e $\mathbb{V}[S_n^2] = \frac{2\sigma^4}{n-1}$,
  - $\widehat{\sigma}_n^2 \sim \mathrm{Gamma}\left(\frac{n-1}{2}, \texttt{rate} = \frac{n}{2\sigma^2}\right)$, $\mathbb{E}[\widehat{\sigma}_n^2] = \frac{n-1}{n}\sigma^2$ e $\mathbb{V}[\widehat{\sigma}_n^2] = \frac{2(n-1)\sigma^4}{n^2}$,
  - $\frac{(n-1)S_n^2}{\sigma^2} = \frac{n\widehat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1}$

- $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ iid $\Rightarrow$
  - $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim \mathrm{N}(0,1)$
  - $\frac{n(\overline{X}_n - \mu)^2}{\sigma^2} \sim \chi^2_1$

- t di Student
  - $U \sim \mathrm{N}(0,1)$ e $V \sim \chi^2_p$ **indipendenti** $\Longrightarrow T = \frac{U}{\sqrt{\frac{V}{p}}} = \frac{\sqrt{p}U}{\sqrt{V}} \sim t_p$

  (t di Student con p gradi di libertà)
  - $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ iid $\Longrightarrow U = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim \mathrm{N}(0,1)$ e $V = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$
  - U e V indipendenti
  - $T = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}$

5. **Campionamento da popolazioni gamma**

- $X \sim \mathrm{Gamma}(\alpha, \texttt{rate} = \beta)$ e $a > 0 \quad \Longrightarrow \quad aX \sim \mathrm{Gamma}(\alpha, \texttt{rate} = \beta/a)$.

- $X \sim \mathrm{Gamma}(\alpha, \texttt{scale} = \beta)$ e $a > 0 \quad \Longrightarrow \quad aX \sim \mathrm{Gamma}(\alpha, \texttt{scale} = a\beta)$.

- Relazioni

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

- Integrale gamma: $\int_0^\infty x^{a-1}e^{-bx}dx = \frac{\Gamma(a)}{b^a}$, con $a, b > 0$

- Variabile aleatoria gamma inversa[1]

  - $X \sim \text{Ga}(\alpha, \texttt{rate} = \beta) \implies Y = \frac{1}{X} \sim \text{IGa}(\alpha, \texttt{rate} = \beta)$
  - $f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{y^{\alpha+1}} e^{-\frac{\beta}{y}}, \quad y \geq 0$
  - $\mathbb{E}[Y] = \frac{\beta}{\alpha-1}$, $\mathbb{V}[Y] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ (esistono rispettivamente solo se $\alpha > 1$ e $\alpha > 2$)
  - Integrale gamma inversa: $\int_0^\infty \frac{1}{x^{a+1}} e^{-\frac{\beta}{x}} dx = \frac{\Gamma(a)}{b^a}$, con $a, b > 0$
  - Tutto analogo per il caso `scale`

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 Unif$[0, \theta] \implies$
  - $F_X(x) = \frac{x}{\theta} I_{[0,\theta]}(x)$
  - $F_{X_{(n)}}(x) = \big[F_X(x)\big]^n = \big[\frac{x}{\theta}\big]^n I_{[0,\theta]}(x)$
  - $f_{X_{(n)}}(x) = n \left[\frac{x}{\theta}\right]^{n-1} \frac{1}{\theta} I_{[0,\theta]}(x) = n \frac{x^{n-1}}{\theta^n} I_{[0,\theta]}(x)$
  - $\mathbb{E}(X_{(n)}) = \frac{n}{n+1}\theta$
  - $\mathbb{E}(X_{(n)}^2) = \frac{n}{n+2}\theta^2$
  - $\mathbb{V}(X_{(n)}) = \frac{n\theta^2}{(n+1)^2(n+2)}$

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

`Risultato generale` (conseguenza del teorema del limite centrale)

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

`Esempi notevoli`

- $X_1, \ldots, X_n \sim \text{Ber}(\theta)$ iid $\implies$
  - $\overline{X}_n \dot\sim \text{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$
  - $Y_n = \sum_{i=1}^n X_i \dot\sim \text{N}\left(n\theta, n\theta(1-\theta)\right)$

[1] Ricorda che se X è una v.a. assolutamente continua con funzione di densità $f_X(\cdot)$ e se $Y = g(X)$ con g invertibile, allora

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy} g^{-1}(y)\right|,$$

dove $g^{-1}(\cdot)$ indica la funzione inversa di $g(\cdot)$.

- $X_1, \ldots, X_n \sim \text{Pois}(\theta)$ iid $\Longrightarrow$
  - $\overline{X}_n \dot{\sim} \text{N}\left(\theta, \frac{\theta}{n}\right)$
  - $Y_n \dot{\sim} \text{N}\left(n\theta, n\theta\right)$
- $X_1, \ldots, X_n \sim \text{Exp}(\theta)$ iid $\Longrightarrow$
  - $\overline{X}_n \dot{\sim} \text{N}\left(\theta, \frac{\theta^2}{n}\right)$
  - $Y_n \dot{\sim} \text{N}\left(n\theta, n\theta^2\right)$

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

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