

Probabilità e Statistica per l'Informatica

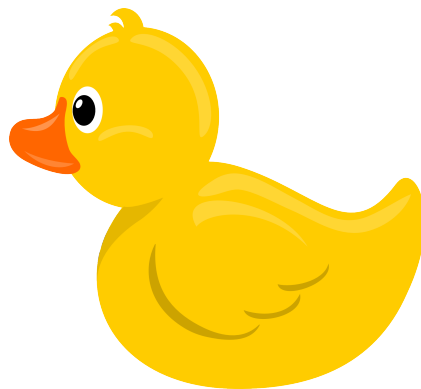
Ver 1.0

Falbo Andrea

A.A 2022/2023

Prof. Caravenna, Masiero

 LilQuacky



Indice

1	Teoremi di Convergenza	2
1.1	Teoria	2
1.1.1	Campione Aleatorio Casuale	2
1.1.2	Teorema del Limite Centrale	4
1.1.3	Correzioni di Continuità	4
1.2	Pratica	5
1.2.1	Formule Utili	5
1.2.2	Esercizi	7

Capitolo 1

Teoremi di Convergenza

1.1 Teoria

1.1.1 Campione Aleatorio Casuale

Il modello probabilistico fondamentale per la statistica è la successione X_1, \dots, X_n di variabili aleatorie **indipendenti ed identicamente distribuite** (v.a i.i.d.)

Definizione: Chiameremo **campione aleatorio casuale** di ampiezza n le X_1, \dots, X_n osservabili.

Osservazione: Uno dei pochi casi in cui è facile conoscere la distribuzione del campione aleatorio è il caso del **campione gaussiano** $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ in quanto possiamo calcolare la media campionaria \bar{X}_N con la tavola della distribuzione.

Obiettivo: Per ottenere il nostro obiettivo, ovvero quello di conoscere la distribuzione di un campione aleatorio X_1, \dots, X_n , dovremo ricondurre il campione aleatorio ad una gaussiano $N(0, 1)$

Definizione: Per standardizzare un campione aleatorio X_1, \dots, X_n dovremo prendere la somma di tale campione, sottrarre la sua media e dividere per la radice della varianza:

$$\frac{X_1 + \dots + X_n - E[X]}{\sqrt{Var[X]}}$$

Osservazioni: Guardiamo ora le distribuzioni di X_i, \dots, X_n per campioni non normali. Avremo che le densità si riconducono a quella di una normale *all'aumentare di n* .

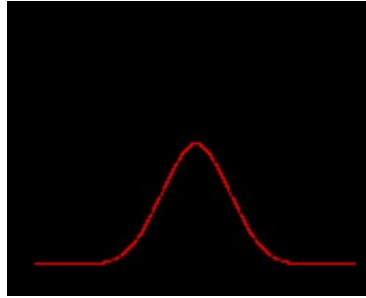


Figura 1.1: Uniforme Continua di parametro $U(0,1)$ con $n=16$

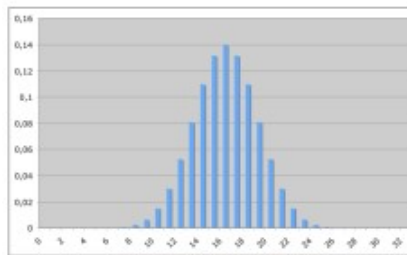


Figura 1.2: Bernoulli di parametro $Be(1/2)$ con $n=32$

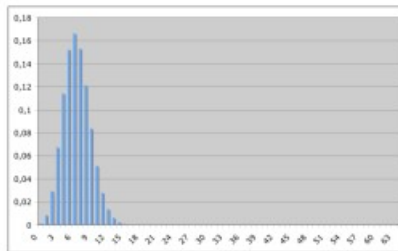


Figura 1.3: Bernoulli di parametro $Be(1/10)$ con $n=64$

1.1.2 Teorema del Limite Centrale

Definizione: Il **Teorema del Limite Centrale** (TLC) afferma che la somma o la media di un grande numero di v.a. i.i.d è approssimativamente normale per n grande ($n \geq 30$). Siano $E[X_i] = \mu$ e $Var[X_i] = \sigma^2$ allora

$$\text{Media: } \mathbb{P} \left(\frac{\overline{X_N} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x \right) \sim P(Z \leq x)$$

$$\text{Somma: } \mathbb{P} \left(\frac{X_i + \dots + X_n - \mu}{\sigma} \leq x \right) \sim P(Z \leq x)$$

1.1.3 Correzioni di Continuità

Definizione: La correzione di continuità si applica quando si standardizza una distribuzione discreta ad una normale. Afferma che bisogna ampliare di $1/2$ gli estremi dell'intervallo per ottenere un valore ben approssimato

Definizione: La correzione di continuità di una **binomiale** afferma che dato se $np \geq 5$ e $n(1-p) \geq 5$ allora la sua correzione è

$$X \sim \text{Bin}(np, np(1-p))$$

Definizione: Se non ci troviamo nel caso di prima, quindi $np \leq 5$ e $n(1-p) \leq 5$ allora approssimiamo con la correzione di continuità di una **Poisson**:

$$X \sim \text{Pois}(n(1-p))$$

1.2 Pratica

1.2.1 Formule Utili

Distribuzioni notevoli discrete					
<i>Distribuzione</i>	$X(\Omega)$	$p_X(k)$ per $k \in X(\Omega)$	$E[X]$	$\text{Var}[X]$	
Bernoulli					
Be(p) $p \in [0, 1]$	$\{0, 1\}$	$\begin{cases} p & \text{se } k = 1 \\ 1 - p & \text{se } k = 0 \end{cases}$	p	$p(1 - p)$	
Binomiale					
Bin(n, p) $n \in \{1, 2, \dots\}$ $p \in [0, 1]$	$\{0, 1, \dots, n\}$	$\binom{n}{k} p^k (1 - p)^{n-k}$	np	$np(1 - p)$	
Poisson					
Pois(λ) $\lambda \in (0, \infty)$	$\mathbb{N}_0 = \{0, 1, \dots\}$	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ	
Geometrica					
Geo(p) $p \in (0, 1]$	$\mathbb{N} = \{1, 2, \dots\}$	$p(1 - p)^{k-1}$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$	
Distribuzioni notevoli assolutamente continue					
<i>Distribuzione</i>	$X(\Omega)$	$f_X(x)$ per $x \in X(\Omega)$	$F_X(x)$	$E[X]$	$\text{Var}[X]$
Uniforme continua					
U(a, b) $a, b \in \mathbb{R}$ con $a < b$	$[a, b]$	$\frac{1}{b - a}$	$\frac{x - a}{b - a}$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Esponenziale					
Exp(λ) $\lambda \in (0, \infty)$	$[0, \infty)$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale					
N(μ, σ^2) $\mu \in \mathbb{R}$ $\sigma \in (0, \infty)$	$(-\infty, +\infty)$	$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$	$\Phi(x)$	μ	σ^2

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	0.99995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	0.99997									
5.0	0.9999997									
6.0	0.9999999990									

Standardizzazione: $\frac{X_i + \dots + X_n - E[X]}{\sqrt{Var[X]}}$

TLC:

$$\text{Media: } \mathbb{P} \left(\frac{\overline{X_N} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x \right) \sim P(Z \leq x)$$

$$\text{Somma: } \mathbb{P} \left(\frac{X_i + \dots + X_n - \mu}{\sigma} \leq x \right) \sim P(Z \leq x)$$

CC: Riduzione di $1/2$ del dominio per approssimare V.A. discrete

CC Binomiale: se $np \geq 5$ e $n(1-p) \geq 5$ allora: $X \sim Bin(np, np(1-p))$

CC Poisson: se $np \leq 5$ e $n(1-p) \leq 5$ allora: $X \sim Pois(n(1-p))$

1.2.2 Esercizi

Esercizio 1: TLC con assolutamente continua

Traccia: Una lampada ha un tempo di vita che segue una legge esponenziale di media 10 giorni. Non appena la lampada smette di funzionare, viene sostituita con una nuova.

- Qual è la probabilità che 40 lampade siano sufficienti per un anno?
- Qual è il numero minimo n di lampade comprare affinché la probabilità dell'evento "n lampade siano sufficienti per un anno" sia almeno 0.95?

Soluzione a:

- Ricavo dalla traccia che ho 40 lampade con legge esponenziale di media 10 giorni. Ricavo dalla tabella che se $E[X] = 1/\lambda = 10$, allora $\lambda = 1/10$. Riscrivo come X_1, \dots, X_{40} v.a. i.i.d. $\sim \exp(1/10)$
- La richiesta è $\mathbb{P}(X_1 + \dots + X_{40} > 365)$. Dato che $n = 40$ è abbastanza grande, possiamo stimare questa probabilità con il TLC.
- Dobbiamo standardizzare la somma delle nostre v.a. Per farlo abbiamo bisogno di conoscere valore medio $E[X] = 1/\lambda$ e varianza $Var[X] = 1/\lambda^2$:

- Per la linearità della media, $E[X + Y] = E[X] + E[Y]$:

$$E[X_1 + \dots + X_{40}] = E[X_1] + \dots + E[X_{40}] = 10 + \dots + 10 = 400$$

- Per l'indipendenza delle v.a., $Cov[X+Y] = 0$ quindi $Var[X+Y] = Var[X] + Var[Y]$:

$$Var[X_1 + \dots + X_{40}] = Var[X_1] + \dots + Var[X_{40}] = 100 + \dots + 100 = 4000$$

- Standardizziamo con la formula: $\frac{X_i + \dots + X_n - E[X]}{\sqrt{Var[X]}}$:

$$\mathbb{P}(X_1 + \dots + X_{40} > 365) = \mathbb{P}\left(\frac{X_1 + \dots + X_{40} - 400}{\sqrt{4000}} > \frac{365 - 400}{\sqrt{4000}}\right)$$

- Semplifichiamo e cambiamo $>$ in \leq :

$$= 1 - \mathbb{P}\left(\frac{X_1 + \dots + X_{40} - 400}{\sqrt{4000}} \leq \frac{-35}{20\sqrt{10}}\right)$$

- Essendo $\sim N(0, 1)$ usando il TLC avremo:

$$\simeq 1 - \mathbb{P}\left(Z \leq \frac{-35}{20\sqrt{10}}\right) = 1 - \Phi\left(\frac{-35}{20\sqrt{10}}\right)$$

- Essendo l'argomento negativo diventa:

$$1 - 1 - \Phi\left(\frac{35}{20\sqrt{10}}\right) = \Phi(0.55) = 0.7088$$

5. La probabilità che 40 lampade siano sufficienti per un anno è 71%

Soluzione b:

1. Abbiamo $X_1 + \dots + X_n \sim \exp(1/10)$. L'incognita ora è n . Cerchiamo il minimo n tc $\mathbb{P}(X_1 + \dots + X_n > 365) \geq 0.95$
2. Sicuramente n deve essere > 40 in quanto al punto a. dava probabilità 0.71 e a noi serve 0.95. Essendo n grande, il TLC è applicabile con buona approssimazione.
3. Dobbiamo standardizzare la somma delle nostre v.a. Calcoliamo valore medio e varianza:

- $E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n] = 10n$
- $Var[X_1 + \dots + X_n] = Var[X_1] + \dots + Var[X_n] = 100n$

4. Standardizzo:

$$\mathbb{P}(X_1 + \dots + X_n > 365) = 1 - \mathbb{P}\left(\frac{X_1 + \dots + X_n - 10n}{\sqrt{100n}} \leq \frac{365 - 10n}{\sqrt{10\sqrt{n}}}\right)$$

5. Avrò numeratore negativo perché $n > 40$ quindi per TLC:

$$\Phi\left(\frac{10n - 365}{10\sqrt{n}}\right)$$

6. Calcoliamo il minimo n tc.

$$\Phi\left(\frac{10n - 365}{10\sqrt{n}}\right) \geq 0.95$$

7. Calcolo con la fdr lo z^* tc $\Phi(z^*) = 0.95$ e trovo $\Phi(1.645)$:

$$\begin{aligned}\Phi\left(\frac{10n - 365}{10\sqrt{n}}\right) &\geq \Phi(1.645) = \frac{10n - 365}{10\sqrt{n}} \geq 1.645 = \quad (\sqrt{n} = x) \\ &= 10x^2 - 16.45x - 365 \geq 0 \rightarrow x = 6.92 = n \sim 47.89\end{aligned}$$

8. Il minimo numero di lampadine da compare è 48

Esercizio 2: TLC con discreta + CC

Traccia: Qual è la probabilità di ottenere almeno 29 teste in 50 lanci di una moneta equilibrata?

Soluzione:

1. Definiamo X_1, \dots, X_{50} v.a. i.i.d $\sim Be(1/2)$
2. Cerco il numero di teste nei 50 lanci: $\mathbb{P}(X_1 + \dots + X_{50} \geq 29)$, ovvero $Bin(50, \frac{1}{2})$ con numero di successi $k = 29$
3. Devo usare la CC di una binomiale, dunque devo rispettare due condizioni:
 - (a) $np \geq 5$: $np = 50 \cdot 1/2 = 25 \geq 5$
 - (b) $n(1 - p) \geq 5$: $n(1 - p) = 50 \cdot 1/2 = 25 \geq 5$
4. Essendo entrambe le condizioni confermate, posso approssimare con una binomiale della forma:

$$X \sim Bin(np, np(1 - p)) = Bin(25, 25/2)$$

5. Approssimo con la CC ed ottengo:

$$\mathbb{P}(X_1 + \dots + X_{50} \geq 28.5) = \mathbb{P}\left(\frac{X_1 + \dots + X_{50} - 25}{\sqrt{\frac{25}{2}}}\right)$$

6. Applico il TLC:

$$\mathbb{P}(X_1 + \dots + X_{50} \geq 28.5) \sim 1 - \mathbb{P}\left(Z \leq \frac{7}{\sqrt{50}}\right) = 1 - \Phi(0.99) = 0.1611$$

7. La probabilità di ottenere almeno 29 teste in 50 lanci è del 16%