Lezioni 29/11/2022 **e** 30/11/2022

Variabile normale

La distribuzione normale

Definizione (Distribuzione normale)**.** *Una variabile aleatoria continua si dice avere distribuzione* normale *(o* gaussiana*) di parametri μ e σ^2, con μ e $\sigma > 0$, se la sua densità di probabilità è la funzione:*

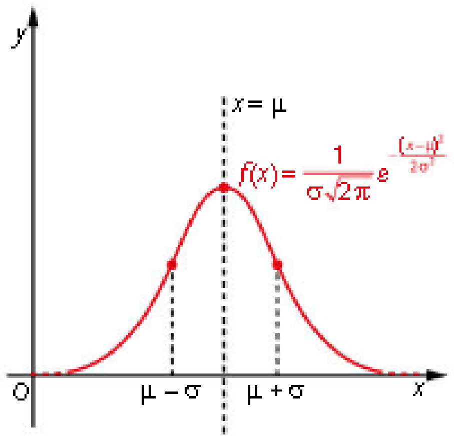
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Per indicare che una variabile aleatoria X è normale di parametri μ e σ^2 si scrive: $X \sim N(\mu, \sigma^2)$. Il fatto che i parametri siano indicati con le lettere μ e σ^2 non è casuale, infatti sussiste il teorema seguente.

Teorema (Media e varianza di una variabile aleatoria normale)**.** *La media e la varianza di una variabile aleatoria X normale di parametri μ e σ^2 sono espresse dalle formule:*

$$E(X) = \mu \qquad e \qquad Var(X) = \sigma^2$$

Il grafico della densità di probabilità di una variabile aleatoria normale, rappresentato nella seguente figura.



Osserviamo che il grafico ha le seguenti caratteristiche:

- è simmetrico rispetto alla retta di equazione $x = \mu$;

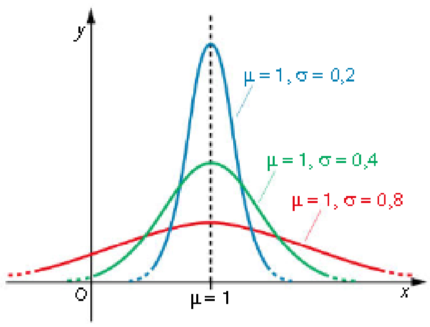
- presenta un massimo nel punto $x = \mu$ (è crescente in $[-\infty, \mu]$ e decrescente in $[\mu, +\infty]$; inoltre $y_{max} = \frac{1}{\sqrt{2\pi}}$);

- presenta due punti di flesso in $x = \mu - \sigma$ e $x = \mu + \sigma$ (ha concavità verso il basso in $[\mu - \sigma, \mu + \sigma]$);

- ha come asintoto l'asse x;

- quanto più σ è piccolo, tanto più il grafico apparirà "appuntito" (la maggioranzadei valori di X saranno infatti vicini a μ) e le "code" si avvicineranno rapidamente all'asse x ($\lim_{x\to-\infty} f(x) = 0$ e $\lim_{x\to+\infty} f(x) = 0$);

- quanto più σ è grande, tanto più il grafico apparirà "smussato" e le "code" si avvicineranno più lentamente all'asse x (vedi la seguente figura).



La distribuzione Normale non descrive in realtà una sola distribuzione, ma piuttosto una famiglia di distribuzioni, tutte con la stessa forma a campana, ma caratterizzate da media e varianza diverse.

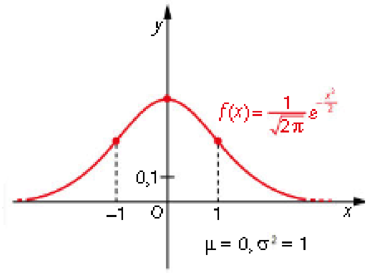
Una variabile aleatoria normale avente media $\mu = 0$ e varianza $\sigma^2 = 1$ è detta **normale standard**; la sua densità è

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

ed il suo grafico è rappresentato dalla seguente figura.

Una variabile aleatoria normale standard viene di solito indicata con la lettera Z: $Z \sim N(0, 1)$.

Curiosità: La distribuzione gaussiana è in realtà stata scoperta nel 1733 dal matematico Abraham De Moivre (1667-1754) e successivamente utilizzata nel 1797 da Gauss(1777-1855) nell'ambito della teoria degli errori.



Come calcolare le probabilità di una variabile aleatoria normale standard

Il calcolo delle probabilità di una variabile aleatoria Z normale standard si basa sulla *funzione di ripartizione* della normale standard, convenzionalmente denotata con la lettera Φ.

$$\Phi(z) = P(Z \leq z)^1 = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx.$$

I valori assunti dalla funzione $\Phi(z)$ si possono ottenere, data la complessità dei calcoli necessari a ricavarli (bisogna risolvere un integrale indefinito), tramite le funzioni predefinite messe a disposizione dai fogli elettronici e dai software di calcolo, oppure facendo riferimento alla tabella riportata nella pagina seguente, che ne elenca i valori per $0 \leq z \leq 3,99$.

[1]Dato che Z è una variabile continua $P(Z \leq z) = P(Z < z)$.

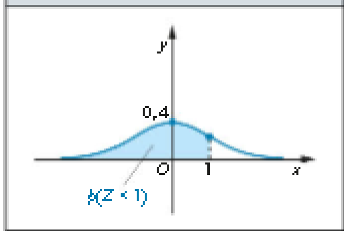
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	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

La prima colonna della tavola riporta i valori di z con la prima cifra decimale; la prima riga riporta la seconda cifra decimale dei valori di z. All'incrocio della riga e della colonna che identificano un prefissato valore di z con due cifre decimali si legge il corrispondente valore di $\Phi(z)$. Per esempio, all'incrocio della riga e della colonna evidenziate si legge il valore di $\Phi(1,24) = 0,8925$ (vedi la seguente figura).

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

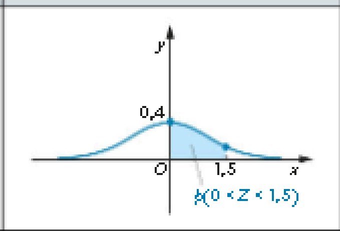
I valori per $z < 0$ si possono ricavare utilizzando le proprietà di simmetria della densità normale standard ($P(Z \geq z) = P(Z \leq -z)$) e la proprietà $P(-\infty \leq Z \leq +\infty) = 1$, come è messo in evidenza nei prossimi esempi.

a. $p(Z < 1)$



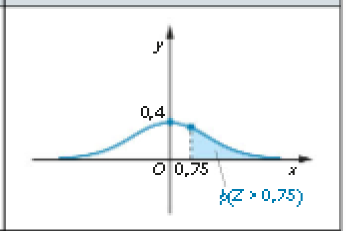
$p(Z < 1) =$
$= \Phi(1) = 0{,}84134$ — Tavole

b. $p(0 < Z < 1{,}5)$



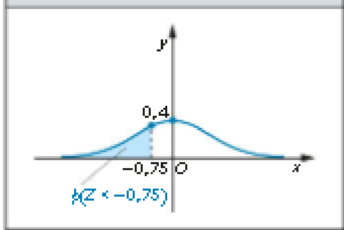
$p(0 < Z < 1{,}5) =$
$= p(Z < 1{,}5) - p(Z \leq 0) =$
$= \Phi(1{,}5) - \Phi(0) =$ — Tavole
$= 0{,}93319 - 0{,}5 =$
$= 0{,}43319$

c. $p(Z > 0{,}75)$



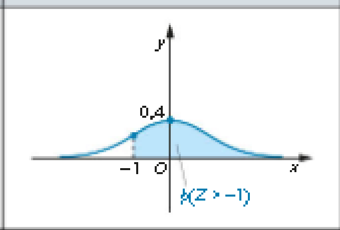
$p(Z > 0{,}75) =$
$= 1 - p(Z \leq 0{,}75) =$
Passaggio al complementare
$= 1 - \Phi(0{,}75) =$
$= 1 - 0{,}77337 = 0{,}22663$ — Tavole

d. $p(Z < -0{,}75)$



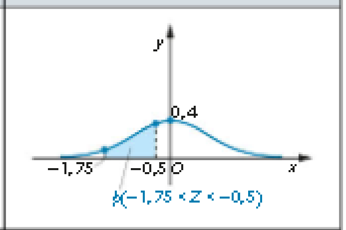
$p(Z < -0{,}75) =$ — Simmetria
$= p(Z > 0{,}75) =$ — Vedi punto c
$= 1 - p(Z \leq 0{,}75) =$
$= 1 - \Phi(0{,}75) = 0{,}22663$

e. $p(Z > -1)$



$p(Z > -1) =$ — Simmetria
$= p(Z < 1) =$
$= \Phi(1) = 0{,}84134$ — Tavole

f. $p(-1{,}75 < Z < -0{,}5)$



$p(-1{,}75 < Z < -0{,}5) =$ — Simmetria
$= p(0{,}5 < Z < 1{,}75) =$
$= p(Z < 1{,}75) - p(Z \leq 0{,}5) =$
$= \Phi(1{,}75) - \Phi(0{,}5) =$
$= 0{,}95994 - 0{,}69146 = 0{,}26848$

Come per i precedenti esempi, si può dimostrare che più in generale vale il seguente teorema.

Teorema. *Per ogni* $z > 0$*,*

(a) $P(Z < -z) = 1 - \Phi(z)$*;*

(b) $P(Z > z) = 1 - \Phi(z)$*;*

(c) $P(-z \leq Z \leq z) = 2\Phi(z) - 1$*.*

Dimostrazione. *(a)* $P(Z \leq -z) = P(Z \geq z)$ *(per le proprietà di simmetria della distribuzione normale)=* $1 - P(Z \leq z)$ *(perchè* $Z \geq z$ *è il complementare di* $Z \leq z$*) = 1 -* $\Phi(z)$*.*

(b) $P(Z \geq z) = 1 - P(Z \leq z)$ *(perchè* $Z \geq z$ *è il complementare di* $Z \leq z$*) = 1 -* $\Phi(z)$*.*

(c) $P(-z \leq Z \leq z)$ *è l'area del trapezoide la cui base ha estremi* $-z$ *e* z*. Quest'area è uguale alla differenza tra* $\Phi(z)$ *e* $\Phi(-z)$*, cioè* $P(-z \leq Z \leq z) = \Phi(z) - \Phi(-z) = \Phi(z) - (1 - \Phi(z)) = 2\Phi(z) - 1$*.*

Si può verificare, come per il caso (c), che $P(z_1 \leq Z \leq z_2) = \Phi(z_2) - \Phi(z_1)$ quando $z_1 < z_2$.

Come calcolare le probabilità di una variabile aleatoria normale di parametri qualsiasi, utilizzando le probabilità della normale standard

Il calcolo delle probabilità di una variabile aleatoria normale di parametri qualsiasi può sempre essere riportato al calcolo delle probabilità di una normale standard. Infatti, se $X \sim N(\mu, \sigma^2)$, si dimostra che la variabile aleatoria Z così definita:

$$Z = \frac{X - \mu}{\sigma}$$

è normale standard. Dunque,

$$P(X \leq x) = P(X < x) = \Phi(\frac{x - \mu}{\sigma}).$$

È particolarmente interessante calcolare la probabilità dei cosiddetti **intervalli tipici** della normale, ossia dei tre intervalli:

$$[\mu - \sigma, \mu + \sigma] \quad [\mu - 2\sigma, \mu + 2\sigma] \quad [\mu - 3\sigma, \mu + 3\sigma].$$

Iniziamo a fare i calcoli nel caso del primo intervallo:

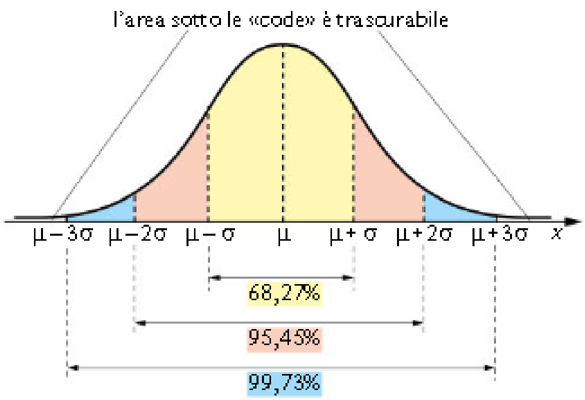
$P(\mu - \sigma \leq X \leq \mu + \sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{\mu + \sigma - \mu}{\sigma}\right) = P(-1 \leq Z \leq 1) = 2\Phi(z) - 1 = 0,68268.$

Analogamente si trova che:

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2\Phi(2) - 1 = 0.9545;$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2\Phi(3) - 1 = 0.9973.$$

Ciò significa che l'area sottesa al grafico della densità normale è quasi interamente contenuta nell'intervallo $[\mu - 3\sigma, \mu + 3\sigma]$, mentre l'area all'esterno di tale intervallo, ossia l'area sotto le code, è trascurabile, essendo meno dello $0,27\%$: è questo il motivo per cui di solito le tavole della normale standard non riportano valori superiori a 3,99 (nel caso della normale standard gli intervalli tipici sono $[-1,1]$,$[-2,2]$ e $[-3,3]$).



Considerando che la variabile normale standard assume il $99,73\%$ dei valori in [-3,3], possiamo calcolare, senza ricorrere ai calcoli o alla tabella, le seguenti probabilità: $P(Z < -8) = 0$, $P(Z \leq 8) = 1$, $P(Z \geq -8) = 1$, $P(Z \geq 8) = 0$ e $P(-8 \leq Z \leq 8) = 1$.

0.0.1 A quali fenomeni si applica il modello normale

Le variabili aleatorie normali sono il modello adatto a interpretare una vasta gammadi fenomeni; in particolare, il modello normale funziona bene quando la variabile aleatoria in esame tende ad assumere un valore prevalente μ o valori per lo più vicinia μ e lo scostamento da μ dipende dalla somma di numerosi fattori aleatori indipendenti.

Esempi tipici di variabili aleatorie che rientrano in queste ipotesi, per le qualisi utilizza il modello normale, sono le seguenti:

- la misura di una grandezza con uno strumento; μ è l'effettiva misura della grandezza mentre il valore ottenuto con lo strumento differisce dalla misura *vera* a causa di tanti piccoli errori, alcuni sistematici (dovuti per esempio allo strumento di misura) e altri casuali e indipendenti tra loro (dovuti per esempio a piccole variazioni della temperatura o della pressione);

- l'effettiva dimensione di un oggetto prodotto da una macchina utensile (μ è il valore nominale della dimensione dell'oggetto mentre il valore effettivo, come nel caso precedente, ne differisce a causa di tanti piccoli errori);

- la misura di caratteristiche quantitative di una popolazione (quali per esempio l'altezza o il peso) che presentino variazioni casuali (ma contenute) intorno a una media.

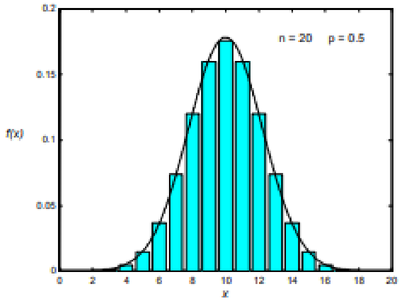
Approssimazione della binomiale con la normale

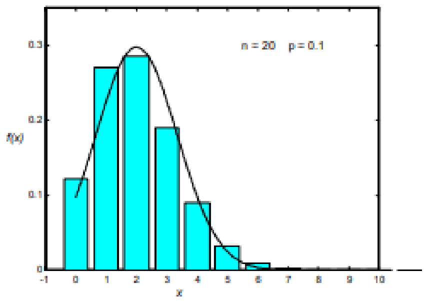
Sia X la variabile aleatoria che fornisce il numero di teste in 1000 lanci di una moneta regolare, si vuole calcolare $P(X \leq 500)$. Osserviamo che $X \sim Bin(1000, 0.5)$ e il $P(X \leq 500) = P(X = 0) + P(X = 1) + \ldots + P(X = 500)$. In generale, quando il numero n delle prove è grande, il calcolo con la distribuzione binomiale è molto lungo. In tal caso è possibile utilizzare la distribuzione normale per approssimare la distribuzione binomiale. Si può dimostrare che, quando n **è grande** e **p è vicino a 0.5**, la distribuzione binomiale della variabile aleatoria X può essere approssimata da una distribuzione normale con media pn e varianza $np(1-p)$. Notiamo che pn e $np(1-p)$ sono la media e la varianza di $X \sim Bin(n, p)$.

L'approssimazione migliora al crescere di n e per $n \to +\infty$ le due distribuzioni coincidono.

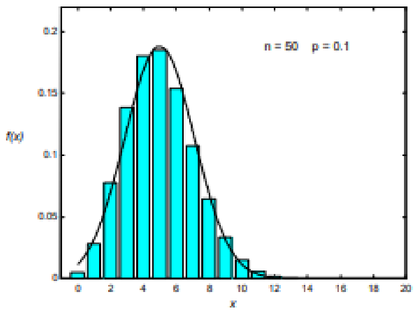
Nella prima figura, per illustrare l'approssimazione fra la distribuzione binomiale e la normale, sono riportati il grafico della distribuzione binomiale per $n = 20$ e $p = 0.5$ e il grafico della distribuzione normale avente valor medio $\mu = np = 10$ e varianza $\sigma^2 = np(1-p) = 5$. L'approssimazione in questo caso risulta buona.

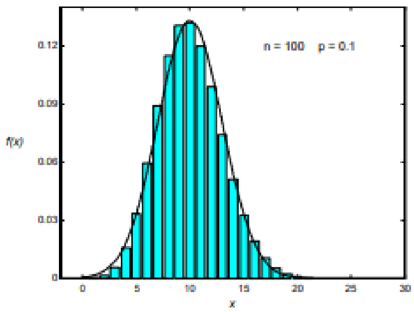
Nella seconda figura, invece, si illustra un caso in cui l'approssimazione della binomiale con la normale non è altrettanto buona ($n = 20$ e $p = 0.1$).





L'approssimazione migliora nei casi seguenti, in cui, malgrado sia p = 0.1, n è più grande.





Per poter usare correttamente la distribuzione normale, che è continua, per approssimare la distribuzione di una variabile aleatoria discreta occorre effettuare la **correzione di continuità**. Essa consiste nel considerare un intervallo leggermente più grande di quello di partenza, per tener conto anche della probabilità negli estremi dell'intervallo (osserviamo che $P(2 \leq X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4)$, ma approssimando $P(2 \leq X \leq 4)$ con una variabile continua commettiamo l'errore di escludere la probabilità negli estremi di $[2, 4]$, perchè se X è una variabile continua allora $P(2 \leq X \leq 4) = P(2 < X < 4)$). Dunque, se X è una variabile aleatoria con distribuzione binomiale di parametri n e p, la probabilità $P(a \leq X \leq b)$ che X assuma valori compresi fra a e b, viene approssimata con il valore della probabilità che la variabile aleatoria normale con media $\mu = np$ e varianza $\sigma^2 = np(1-p)$ assuma valori compresi tra $(a-\frac{1}{2}, b+\frac{1}{2})$. Nel caso particolare in cui $a = b$, la probabilità binomiale $P(X = a)$ viene approssimata con il valore della probabilità $P(a-\frac{1}{2} \leq X \leq a+\frac{1}{2})$ calcolata con la distribuzione normale. Ovviamente $P(X \leq a)$ e $P(X \geq a)$ vengono rispettivamente approssimati con $P(X \leq a + \frac{1}{2})$ e $P(X \geq a - \frac{1}{2})$.

Esercizio 1. *Si effettuano 500 lanci di una moneta regolare, calcola la probabilità che il numero di teste sia minore o uguale di 150.*

Svolgimento. *Dobbiamo calcolare $P(X \leq 150)$, dove $X \sim Bin(500, 0.5)$. Risolvendo il problema utilizzano la distribuzione binomiale, dovremmo calcolare 151 probabilità e poi sommarle: $P(X \leq 150) = P(X = 0) + P(X = 1) + \ldots + P(X = 150)$. E' dunque più facile approssimare X con una variabile normale di parametri $np = 250$ e $np(1-p) = 125$ ($X \sim N(250, 125)$).*

Prima di calcolare $P(X \leq 150)$ con la distribuzione normale effettuiamo la correzione di continuità: $P(X \leq 150 + 0.5) = 150.5$.

$P(X \leq 150.5) = P(Z \leq \frac{150.5 - 250}{\sqrt{125}})$ *(passando alla normale standard)*$=$ $P(Z \leq -8.9) \sim 0$.

Osservazione: Quando n è grande, la binomiale $X \sim Bin(n, p)$ può essere approssimata con

- la variabile normale di parametri np e $np(1-p)$, se p è vicino a $\frac{1}{2}$;
- la variabile di Poisson di parametro np se p è vicino a 0 o 1 Ricordiamo che l'approssimazione della binomiale con la variabile di Poisson si utilizza quando $p \sim 0$. Se $p \sim 1$, si può comunque ricorrere alla variabile di Poisson scrivendo l'evento di cui vogliamo calcolare la probabilità in termini dell'insuccesso. Ad esempio, se X conta in numero di teste nei lanci ripetuti di una moneta e $X \sim Bin(500, 0.9)$, allora $P(X = 150) = P(Y = 350)$, dove Y è il numero di croci e $Y \sim Bin(500, 0.1)$. Possiamo calcolare $P(Y = 350)$ approssimando Y con una variabile di Poisson perchè $p \sim 0$.