# Approssimazione normale alla binomiale

# Calcolo di probabilità binomiali

Numero di successi X in *n* prove indipendenti e identiche con probabilità di successo *p* in ogni prova

Valore atteso e varianza:

$$\boxed{E(X) = \mu = np}$$

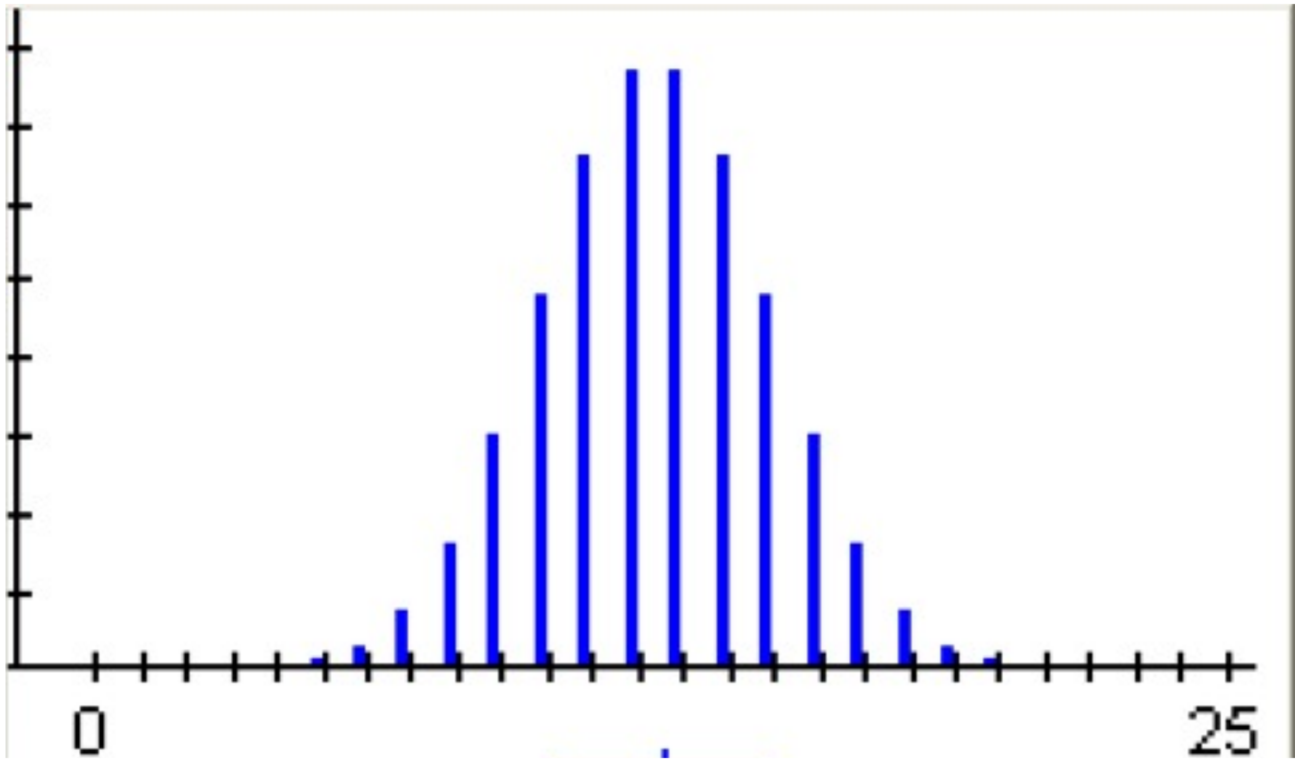
$$\boxed{Var(X) = \sigma^2 = np(1-p)}$$

Quando *n* è grande il calcolo delle probabilità cumulate è complesso: X $\sim$ Bin(*n*=50, *p*)

P(X≤25) = P(X=0) + P(X=1) + … + P(X=25)

# Approssimare la Binomiale con la Normale

Quando *la varianza è* > 9 la Normale è una buona approssimazione della binomiale




In tal caso le probabilità cumulate della **X~Bin(n, p)** sono approssimate da quelle della normale **Y~N(np, np(1-p))**

$$P(X \le a) \simeq P(Y \le a) = P\left(Z \le \frac{a - np}{\sqrt{np(1-p)}}\right)$$

# Esempio

40% dei cittadini sono favorevoli all'operato del sindaco. Qual è la probabilità che, in un campione di $n$ = 200, il numero di favorevoli sia compreso tra 76 e 80 (ovvero la percentuale di favorevoli sia compresa tra 38% e 40%)?

E($X$) = $np$ = 200(0.40) = 80

Var($X$) = $np$(1 – $p$) = 200(0.40)(1 – 0.40) = 48

$$
\begin{aligned}
P(76 < X < 80) &\simeq P\left(\frac{76-80}{\sqrt{48}} < Z < \frac{80-80}{\sqrt{48}}\right) \\
&= P(-0.58 < Z < 0) \\
&= F(0) - F(-0.58) \\
&= 0.5000 - 0.2810 = 0.2190
\end{aligned}
$$

# Combinazioni lineari di più variabili

# Somma di variabili aleatorie

Siano date k variabili aleatorie $X_1, X_2, \ldots, X_k$ (discrete o continue) con medie $\mu_1, \mu_2, \ldots \mu_k$ e varianze $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$.

**Esempio**: i tempi di 4 atleti della squadra di una corsa a staffetta sono $X_1, X_2, X_3, X_4$ . Il tempo totale della squadra è $W = X_1 + X_2 + X_3 + X_4$

Allora, **la media della loro somma è la somma delle loro medie**

$$\boxed{E(X_1 + X_2 + \cdots + X_k) = \mu_1 + \mu_2 + \cdots + \mu_k}$$

# Somma di variabili aleatorie

Siano date k variabili aleatorie $X_1, X_2, \ldots, X_k$ (discrete o continue) con medie $\mu_1, \mu_2, \ldots \mu_k$ e varianze $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$.

Allora, **la varianza della loro somma è la somma delle loro varianze più il doppio delle covarianze tra coppie di variabili**

$$\operatorname{var}(X_1 + \cdots + X_k) = \sigma_1^2 + \cdots + \sigma_k^2 + \sum_{i<j} 2\operatorname{cov}(X_i, X_j)$$

**Se le variabili sono incorrelate a 2 a 2 allora la varianza della somma è la somma delle varianze**

# Combinazione lineare di 2 variabili

Date due variabili aleatorie, X e Y,
W = *a*X + *b*Y si chiama **combinazione lineare**

La media di W è

$$\mu_W = E[W] = E[aX + bY] = a\mu_X + b\mu_Y$$

La varianza di W è

$$\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X, Y)$$

**Se entrambe X e Y sono distribuite normalmente allora anche la combinazione lineare, W, è distribuita normalmente**

# Differenza tra 2 variabili aleatorie

La differenza X - Y è un caso speciale di combinazione lineare

La media della differenza è la differenza delle medie:

$$\boxed{E(X-Y)=\mu_X-\mu_Y}$$

Se la covarianza tra X e Y è 0, allora la varianza della differenza è la somma delle varianze:

$$\boxed{Var(X-Y)=\sigma_X^2+\sigma_Y^2}$$

Se la covarianza tra X e Y non è 0, allora la varianza della differenza include anche la covarianza:

$$\boxed{Var(X-Y)=\sigma_X^2+\sigma_Y^2-2Cov(X,Y)}$$

# Esempio

Due mansioni devono essere eseguite dallo stesso lavoratore

$X$ = tempo (min) necessario per completare mansione 1;
$\mu_X = 20$, $\sigma_X = 5$
$Y$ = tempo (min) necessario per completare mansione 2;
$\mu_Y = 30$, $\sigma_Y = 8$
$X$ e $Y$ sono distribuite normalmente e sono incorrelate

Quali sono la media e la deviazione std del tempo necessario per completare entrambe le mansioni? Qual è la distribuzione?

# Esempio /2

$\mu_X = 20,\ \sigma_X = 5$ $\mu_Y = 30,\ \sigma_Y = 8$

W=X+Y = **tempo necessario per completare entrambe le mansioni**

E(W) = E(X) + E(Y) = 20 + 30 = 50 min

Poiché cov(X,Y) = 0

var(W) = var(X) + var(Y) = 5*5 + 8*8 = 89

La deviazione std è $\boxed{\sigma_W = \sqrt{89} = 9.434}$

La distribuzione è normale $\boxed{W \sim N(50, 89)}$