Esercitazioni
Probabilità e Variabili Aleatorie

Corso di
Statistica

Domenico De Stefano

Maggio 2017

Sintesi
Distribuzioni di probabilità discrete: la binomiale

Variabili aleatorie discrete

Una variabile aleatoria (o variabile casuale) X si dice *discreta* se essa assume solo valori discreti. X può assumere un numero finito o un'infinità numerabile di valori indicati con x.
La probabilità dei valori di X sono espresse tramite la *funzione di probabilità*

$$f(x) = \Pr(X = x)$$

tale che

$$f(x) \geq 0, \forall x \quad \sum_x f(x) = 1$$

La *funzione di ripartizione* è espressa da

$$F(x) = \Pr(X \leq x) = \sum_{t \leq x} f(t)$$

ed è caratterizzata dalle seguenti proprietà:

1. $\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1;$
2. $F(x)$ non decrescente;
3. $F(x)$ continua a destra: $\lim_{h \to 0^+} F(x + h) = F(x);$
4. $\Pr(a < X \leq b) = F(b) - F(a).$

Il *valore atteso* (media) e la *varianza* della variabile aleatoria discreta X sono dati rispettivamente da:

$$\mu = \mathrm{E}(X) = \sum_x x \Pr(X = x)$$

e

$$\sigma^2 = \mathrm{Var}(X) = E(X - \mu)^2$$

ed inoltre vale $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2$. La varianza di X è nulla solo se X assume un unico valore con probabilità 1 ed è tanto più grande quanto maggiore è la dispersione dei valori di X intorno alla media.

Distribuzione binomiale

Si considerino n prove Bernoulliane indipendenti cioè dove possono verificarsi solo 2 possibili eventi (chiamiamoli 'successo' e 'insuccesso'). Sia p la probabilità dell'evento successo. La variabile aleatoria binomiale è definita come il numero dei successi in n prove Bernoulliane indipendenti. La funzione di probabilità è espressa da

$$f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x},\ x = 0, 1, \ldots, n.$$

La distribuzione dipende dalle costanti n e p i cui valori determinano l'andamento della distribuzione (i cosiddetti parametri). La media e la varianza della distribuzione binomiale sono date rispettivamente da

$$\mathrm{E}(X) = np, \quad \mathrm{Var}(X) = np(1-p).$$

Esercizio 1

Presso uno sportello bancomat del centro di Trieste 4 persone su 5 fanno operazione di prelievo. La banca vorrebbe sapere le seguenti probabilità.

a) Supponendo di estrarre a caso 10 persone (con reimmissione) che si sono recate allo sportello, calcolare la probabilità che il numero totale di persone che ha prelevato sia 9. Indicare che tipo di variabile aleatoria descrive questo tipo di esperimento.

b) Supponendo di estrarre a caso 10 persone (con reimmissione) che si sono recate allo sportello, calcolare la probabilità che tutte e 10 abbiano prelevato.

c) Supponendo di estrarre a caso 10 persone (con reimmissione) che si sono recate allo sportello, calcolare la probabilità che il numero totale di persone che ha prelevato sia 11

d) Supponendo di estrarre a caso 10 persone (con reimmissione) che si sono recate allo sportello, calcolare la probabilità che il numero totale di persone che ha prelevato sia maggiore di 8.

e) Rappresentare graficamente la distribuzione di probabilità della variabile aleatoria X descritta ai punti precedenti

f) Qual è la moda di X? media e varianza?

Parte a)
Sia la variabile aleatoria di interesse X= numero di persone che ha prelevato. Vogliamo sapere qual è la probabilità che X=9 su n=10 prove indipendenti (NB:'estrarre a caso 10

persone'). I possibili eventi per ogni estrazione sono 2 (prelevato oppure non prelevato, incompatibili tra loro). La probabilità di estrarre una persona che ha prelevato è pari a 4/5 = 0.8.
La variabile aleatoria X che conta il numero di persone che hanno prelevato si distribuisce pertanto come una Binomiale di parametri p = 0.8 e n = 10. in simboli $X \sim Bin(p = 0.8, n = 10)$. La banca vuole sapere la seguente probabilità: $P(X = 9)$, cioè bisogna calcolare la funzione di probabilità per x=9 ossia $f(9)$
Sapendo che la funzione di probabilità di una binomiale è $f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$ per $x = 0, 1, \ldots, n$. Allora possiamo facilmente calcolare la probabilità richiesta:

$$f(9) = \frac{10!}{9!(10-9)!}0.8^9(1-0.8)^{10-9} =$$

$$f(9) = \frac{10!}{9!(1)!}0.8^9(0.2)^1 =$$

$$\frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 \times (1)}0.8^9(0.2)^1$$

Semplificando

$$\frac{10}{1}0.8^9(0.2) = 10 \times 0.8^9 \times 0.2 = 10 \times 0.134 \times 0.2 = 0.27$$

Vi è una probabilità di 0.27 che 9 persone su 10 prelevino allo sportello. Cioè $P(X = 9) = f(9) = 0.27$
Es. b)

Siamo sempre nel caso $X \sim Bin(p = 0.8, n = 10)$. Vogliamo sapere $P(X = 10)$ ossia calcolare $f(10)$. Sapendo la formula per la probabilità della binomiale allora:

$$f(10) = \frac{10!}{10!(10-10)!}0.8^{10}(1-0.8)^{10-10} =$$

$$f(10) = \frac{10!}{10!(0)!}0.8^{10}(0.2)^0 =$$

Si ricordi che $0! = 1$, allora avremo che:

$$f(10) = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1(1)}0.8^{10}(0.2)^0 =$$

Semplificando:

$$f(10) = 1 \times 0.8^{10} \times 1 = 0.8^{10} = 0.11$$

La probabilità che tutte e 10 le persone prelevino al bancomat è di 0.11. Ossia $P(X = 10) = f(10) = 0.11$

Es. c)

Il quesito ci chiede $P(X = 11) = f(11)$ che sarebbe la probabilità che 11 persone su 10 abbiano prelevato (non ha senso!!!). Chiaramente $P(X = 11) = f(11) = 0$ perché ricordatevi che la funzione di probabilità di una binomiale con parametri n=10 e p=0.8 è $f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$ **e definita solo per** $x = 0, 1, \ldots, 10$ (altrove è 0).

Es. d)

La probabilità che il numero totale di persone che ha prelevato sia maggiore di 8, ossia $P(X > 8)$ si calcola facendo riferimento alla funzione di ripartizione. Infatti sapendo che:

$$P(X \leq 8) = F(8) = \sum_{t \leq 8} f(t) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) + f(6) + f(7) + f(8)$$

allora

$$P(X > 8) = 1 - F(8) = 1 - \sum_{t \leq 8} f(t) =$$

$$= 1 - [f(0) + f(1) + f(2) + f(3) + f(4) + f(5) + f(6) + f(7) + f(8)] = f(9) + f(10)]$$

pertanto:

$$P(X > 8) = f(9) + f(10) = \frac{10!}{9!(10-9)!}0.8^9(1-0.8)^{10-9} + \frac{10!}{10!(10-10)!}0.8^{10}(1-0.8)^{10-10} =$$

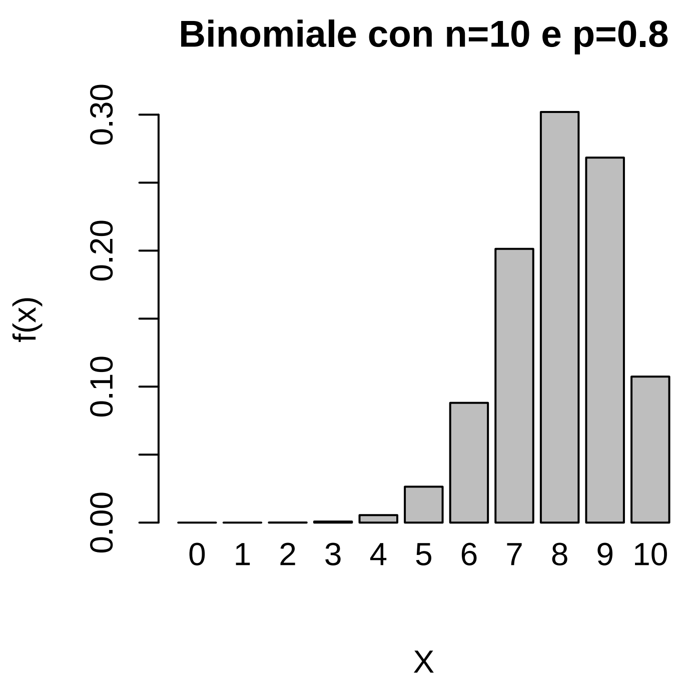
$$= 0.27 + 0.11 = 0.38$$

Es. e)

Per poter rappresentare graficamente la distribuzione di probabilità della variabile aleatoria X occorre calcolare tutte le probabilità $P(X = x)$ per tutti gli $x = 0, 1, \ldots, 10$.
ossia dovreste completare la seguente tabella (abbiamo già calcolato f(9) e f(10)):

X	f(x)
0	f(0)
1	f(1)
2	f(2)
3	f(3)
4	f(4)
5	f(5)
6	f(6)
7	f(7)
8	f(8)
9	f(9) = 0.27
10	f(10) = 0.11

Il seguente diagramma a barre illustra graficamente la distribuzione di probabilità di X:



Es. f)
La moda di X è il valore più probabile ossia: Mo=8.
Media e Varianza possono calcolari utilizzando i parametri della binomiale:

$$E(X) = \mu_X = np = 10 \times 0.8 = 8$$

La varianza è invece uguale a:

$$VAR(X) = \sigma_X^2 = np(1-p) = 10 \times 0.8 \times 0.2 = 1.6$$

pertanto la deviazione standard sarà:

$$SD(X) = \sigma_X = \sqrt{np(1-p)} = \sqrt{10 \times 0.8 \times 0.2} = 1.26$$

Sapreste calcolare anche la mediana?

Esercizio 2

Da un mazzo di 52 carte (13 di picche, 13 di cuori, 13 di fiori e 13 di quadri) ne vengono estratte cinque con reimmissione. Si è interessati alla variabile casuale X che descrive il numero di carte di cuori ottenute nelle estrazioni. Determinare:

a) il valore atteso e la varianza della variabile X

b) la probabilità di estrarre tre carte di cuori.

c) la probabilità di estrarre al massimo tre carte di cuori

d) la probabilità di estrarre tra 2 e 4 carte di cuori

Es. a)
La variabile aleatoria di interesse è X= numero di carte di cuori ottenute. Le estrazioni sono indipendenti e gli eventi sono estrarre carta di cuori vs estrarre carta di altro seme. La probabilità di estrarre una carta di cuori dal mazzo è $\frac{13}{52} = 1/4 = 0.25$. Pertanto possiamo affermare che X si distribuisce come una variabile aleatoria binomiale di parametri $n = 5$ e $p = 0.25$.
Volendo sapere valore atteso e varianza utilizziamo pertanto i parametri per calcolare le seguenti quantità:

$$E(X) = \mu_X = np = 5 \times 0.25 = 1.25$$

La varianza è invece uguale a:

$$VAR(X) = \sigma_X^2 = np(1-p) = 5 \times 0.25 \times 0.75 = 0.94$$

In media nelle n=5 estrazioni ci aspettiamo 1.25 carte di cuori con deviazione standard di 0.968.
Es. b)

Vogliamo la probabilità $P(X = 3)$ ossia $f(3)$. Sapendo che la funzione di probabilità di una binomiale è $f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$ per $x = 0, 1, \ldots, n$. Allora possiamo facilmente calcolare la probabilità richiesta:

$$f(3) = \frac{5!}{3!(5-3)!}0.25^3(1-0.25)^{5-3} =$$

$$f(3) = \frac{5!}{3!(2)!}0.25^3(1-0.25)^2 =$$

$$= \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times (2 \times 1)}0.25^3(0.75)^2 =$$

Semplificando:

$$= \frac{5 \times 4}{(2 \times 1)}0.25^3(0.75)^2 =$$

$$= \frac{20}{2}0.25^3(0.75)^2 = 10 * 0.016 * 0.562 = 0.0879$$

La probabilità di ottenere 3 carte di cuori sulle 5 è di circa il 9%
Es. c)

La probabilità di ottenere al massimo 3 carte corrisponde in simboli alla seguente probabilità $P(X \leq 3)$ ossia $F(3) = f(0) + f(1) + f(2) + f(3)$. Cioè corrisponde alla funzione di ripartizione calcolate in x=3, cioè:

$$P(X \leq 3) = F(3) = \sum_{t \leq 3} f(t) = f(0) + f(1) + f(2) + f(3)$$

pertanto:

$$P(X \leq 3) = \frac{5!}{0!(5-0)!}0.25^0(0.75)^{5-0} + \frac{5!}{1!(5-1)!}0.25^1(0.75)^{5-1} + =$$

$$+\frac{5!}{2!(5-2)!}0.25^2(0.75)^{5-2} + \frac{5!}{3!(2)!}0.25^3(1-0.25)^2 = 0.237 + 0.40 + 0.264 + 0.0879 = 0.9889$$

La probabilità di ottenere al massimo 3 carte di cuori sulle 5 è di circa il 99%.

Variabili aleatorie continue

Si ha una variabile aleatoria *continua* quando essa assume vaolri nel continuo cioè, una infinità non numerabile di eventi elementari.
Una variabile aleatoria X definita nell'intervallo (α, β) [α e β possono anche essere rispettivamente $-\infty$ e ∞] è detta continua se esiste una funzione $f(x)$ detta *funzione di densità di probabilità* di X, tale che

1. $f(x) \geq 0$
2. $\int_{\alpha}^{\beta} f(x)dx = 1$
3. $\int_{a}^{b} f(x)dx = \Pr(a < X < b), \; a < b$

La *funzione di ripartizione* della variabile aleatoria continua X è espressa da

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

ed è caratterizzata dalle seguenti proprietà:

1. $F(x)$ non decrescente e continua;
2. $\Pr(a \leq X \leq b) = F(b) - F(a)$.

Il *valore atteso* e la *varianza* della variabile aleatoria continua X sono espressi da:

$$\mu = \mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)dx$$

purchè l'integrale esista e sia finito, e

$$\sigma^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

ed inoltre vale $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$. La radice quadrata della varianza σ è la *deviazione standard*, misura del grado di dispersione della v.a.

Distribuzione uniforme
Una variabile aleatoria definita nell'intervallo $[a, b]$ ha distribuzione uniforme o rettangolare se la sua funzione di densità è espressa da

$$f(x) = \frac{1}{b-a} \; a \leq x \leq b.$$

La funzione di ripartizione risulta

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > 1 \end{cases}$$

Media e varianza sono date rispettivamente da $\mathrm{E}(X) = (a+b)/2$, $\mathrm{Var}(X) = (a-b)^2/12$.

Distribuzione normale
Una variabile aleatoria continua ha distribuzione normale se la sua funzione di densità è espressa da

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \ \infty < x < -\infty$$

dove μ e σ^2 sono la media e la varianza della variabile aleatoria. La distribuzione normale con media 0 e varianza 1 è detta *normale standardizzata*. La funzione di ripartizione è espressa da

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2} dy.$$

Ne segue che

$$\Pr(a \leq X \leq b) = \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

dove $\Phi()$ indica la funzione di ripartizione della normale standardizzata. Cioè $\Phi()$ indica la probabilità $P(Z \leq z)$ ossia **l'area a sinistra del valore z**.
Inoltre vale la seguente uguaglianza $\Phi(-z) = 1 - \Phi(z)$.
NB: per calcolare qualsiasi probabilità di qualsiasi normale con parametri μ e σ arbitrari ci si riduce sempre al caso di una variabile aleatoria normale standardizzata (e si usano i valori della funzione di ripartizione $\Phi(z)$ tabulati nelle cosiddette tavole della normale standard).

Esercizio 3

Rispondere alle seguenti domande in relazione alla distribuzione normale.

a) Cosa indicano i seguenti simboli: μ, σ, Z, z.

b) Cosa sono le tavole della normale standard?

c) Quanto vale la probabilità $P(Z \leq 0)$ (ossia $\Phi(0)$) sapendo che Z si distribuisce come una normale standard?

d) Quanto vale la probabilità $P(Z \leq 0.63)$ (ossia $\Phi(0.63)$) sapendo che Z si distribuisce come una normale standard?

e) Quanto vale la probabilità $P(Z \leq -0.99)$ (ossia $\Phi(-0.99)$) sapendo che Z si distribuisce come una normale standard?

f) Quanto vale la probabilità $P(Z \geq 1.96)$ (ossia $1 - \Phi(1.96)$) sapendo che Z si distribuisce come una normale standard?

g) Quanto vale la probabilità $P(Z \geq -1.96)$ sapendo che Z si distribuisce come una normale standard?

h) Quanto vale la probabilità $P(-2.3 \leq Z \leq -1.5)$ sapendo che Z si distribuisce come una normale standard?

i) Quanto vale la probabilità $P(-2.3 \leq Z \leq 1.5)$ sapendo che Z si distribuisce come una normale standard?

Es. a)
Quando parliamo di distribuzione normale i simboli μ e σ rappresentano rispettivamente la media e la deviazione standard della normale. Si ricordi che μ può essere un qualsiasi numero reale (dipende dalla variabile che ha forma normale che stiamo considerando) $-\infty < \mu < \infty$ mentre σ un qualsiasi numero reale positivo (ricordate, la varianza non può essere negativa e pertanto nemmeno la deviazione standard) $0 < \sigma < \infty$.
Inoltre μ e σ sono anche i parametri della variabile aleatoria normale, in simboli infatti se una variabile X si distribuisce come una normale scriveremo: $X \sim Nor(\mu, \sigma)$. Questo vuole dire che μ e σ governano la forma della normale: in particolare μ governa la posizione della normale lungo l'asse delle ascisse; σ governa l'ampiezza della curva $f(x)$ e di conseguenza la sua altezza (ricordatevi che l'area sotto la curva deve essere sempre pari a 1 altrimenti $f(x)$ non sarebbe una vera funzione di densità di probabilità).
Nel contesto della normale il simbolo Z indica la variabile aleatoria normale standard e z i possibili valori che questa può assumere che ricordiamo essere $-\infty < z < \infty$. La normale standard è una normale che ha parametri $\mu = 0$ e $\sigma = 1$, cioè $Z \sim Nor(0, 1)$. Essendo costruita in questo modo i valori z della normale standard rappresentano scostamenti dal centro (media) della distribuzione in termini di unità di devizioni standard (es: $z = -2$ vuol dire che quella determinata osservaazione si discosta a sinistra della media di 2 volte la deviazione standard).
Es. b)
La normale standard Z è l'unica normale per la quale i valori della funzione di ripartizione sono tabulati. Questi valori si trovano nelle tavole della normale standard e sono indicati con con $\Phi(z) = P(Z \leq z)$ anzichè col simbolo $F()$ come abbiamo fatto finora. Quindi i valori nelle tavole sono i valori degli integrali che indicano l'area fino al valore

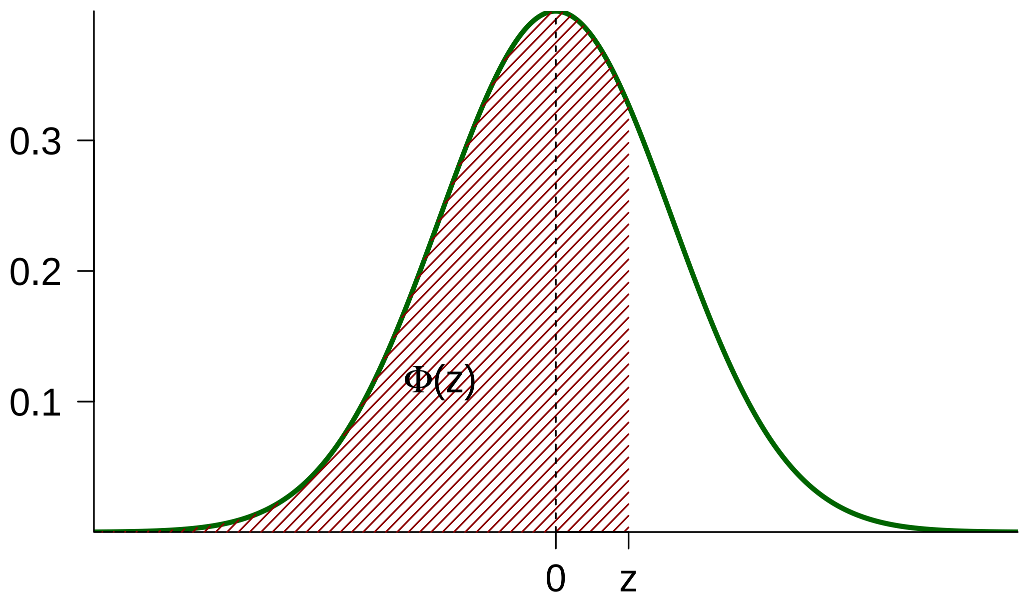
z ossia la probabilità $P(Z \leq z)$.
Si usano nel seguente modo: il valore di z, fino alla prima cifra decimale, è riportato nella prima colonna; la seconda cifra decimale è indicata nella prima riga delle altre colonne e, in corrispondenza di essa, è riportato il valore dell'integrale $\Phi(z)$.
Notate che i valori superiori a 3 non vengono tabulati tutti in quanto la corrispondente della probabilità $\Phi(z)$ comincia ad approssimarsi ad 1.

Es. c)

Dalla proprietà della normale standard sappiamo che il valore $z = 0$ corrisponde alla media che tra l'altro, essendo la distribuzione perfettamente simmetrica, coincide con la moda e la mediana. Proprio quest'ultima ci aiuta a trovare il valore della probabilità $P(Z \leq 0)$.
Infatti sapendo che la mediana è quell'osservazione che lascia alla sua destra e alla sua sinistra il 50% delle osservazioni avremo che $P(Z \leq 0) = \Phi(0) = 0.5$.
E' facile verificarlo anche dalle tavole della normale. Infatti, cerchiamo il volore $z = 0$ nelle tavole: questo si trova in corrispondenza della prima riga (valore $z = 0.0$) e della prima colonna (seconda cifra decimale 0), la cella che incrocia la prima riga e la prima colonna contiene l'integrale desiderato ossia il $\Phi(0)$ (l'area sottesa alla curva in corrispondenza del valore $z = 0$ desiderato).

Es. d)

Possiamo sfruttare il risultato precedente per intuire che $P(Z \leq 0.63) > P(Z \leq 0)$, ossia l'area sottesa fino a 0.63, cioè $\Phi(0.63)$ è maggiore dell'area sottesa fino a 0, cioè $\Phi(0)$ (ovvio no?). Pertanto sicuramente $P(Z \leq 0.63) > 0.5$.
Per i primi esercizi è consigliabile oltre cha fare questo tipo di ragionamenti anche aiutarsi con una rappresentazione grafica della funzione di densità della normale standard. Di seguito è illustrata l'area della normale standard Z fino a 0.63 ossia è illustrato l'integrale $P(Z \leq 0.63) = \Phi(0.63)$:

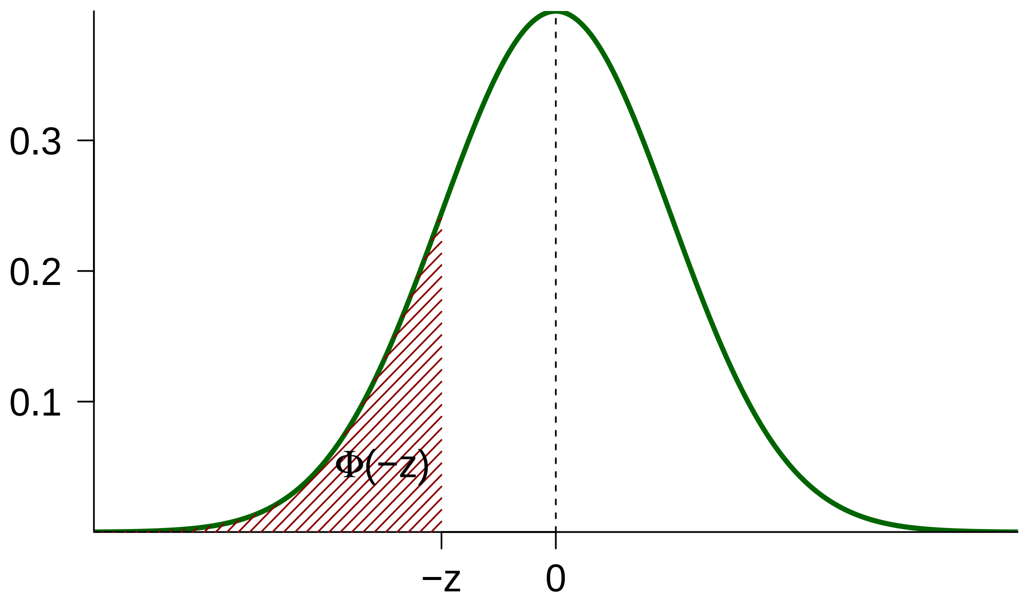


La probabilità richiesta $P(Z \leq 0.63)$ è proprio la funzione di ripartizione valutata in 0.63, ossia $\Phi(0.63)$, pertanto occorrerà semplicemente cercare sulle tavole della normale standard la riga 0.6 e incrociarla con la colonna 0.03. La probabilità di osservare un valore inferiore o uguale a 0.63 è di 0.7357.

Es. e)

Si noti che i valori negativi di z (quelli a inferiori alla media) non sono tabulati ma le proprietà della normale ci consentono agevolmente di individuare le areee sottese ad essi. Si ricordi inoltre che tutta l'area sotto la curva è uguale a 1.

Per individuare la probabilità $P(Z \leq -0.99)$ infatti occorre sfruttare la simmetria della normale e la conseguente proprietà $\Phi(-z) = 1 - \Phi(z)$. Graficamente la probabilità da individuare è:



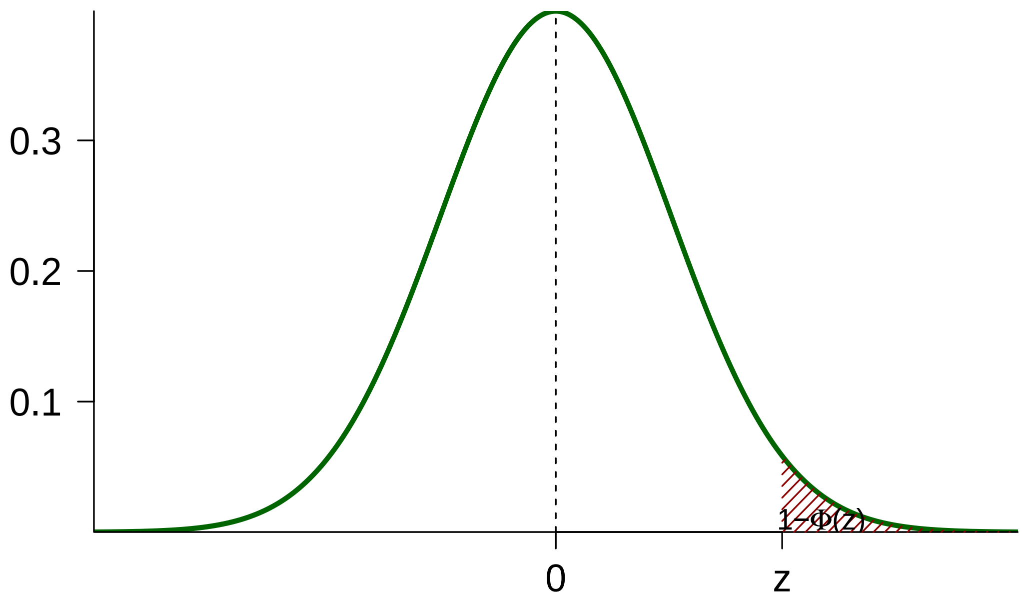
Cercare $P(Z \leq -0.99)$ equivale a valutare la funzione di ripartizione in -0.99 ossia $\Phi(-0.99)$ pertanto secondo la proprietà richiamata sopra quindi:

$$\Phi(-z) = 1 - \Phi(z) \Rightarrow \Phi(-0.99) = 1 - \Phi(0.99)$$

quindi basta trovare $\Phi(0.99)$ sulle tavole e calcolarne il complementare. Dalle tavole incrociando la riga 0.9 con la colonna 0.09 individuiamo $\Phi(0.99) = 0.8389$. Pertanto:

$$\Phi(-0.99) = 1 - 0.8389 = 0.1611$$

La probabilità di osservare un valore inferiore o uguale a -0.99 è di 0.1611.
Es. f)
La probabilità che si cerca in questo esercizio è il complementare della funzione di ripartizione (NB: Attenzione al segno della disuguaglianza!). Infatti cerchiamo $P(Z \geq 1.96)$ ossia $P(Z \geq 1.96) = 1 - P(Z \leq 1.96) = 1 - \Phi(1.96)$ e che corrispnde alla seguente area sotto la curva:



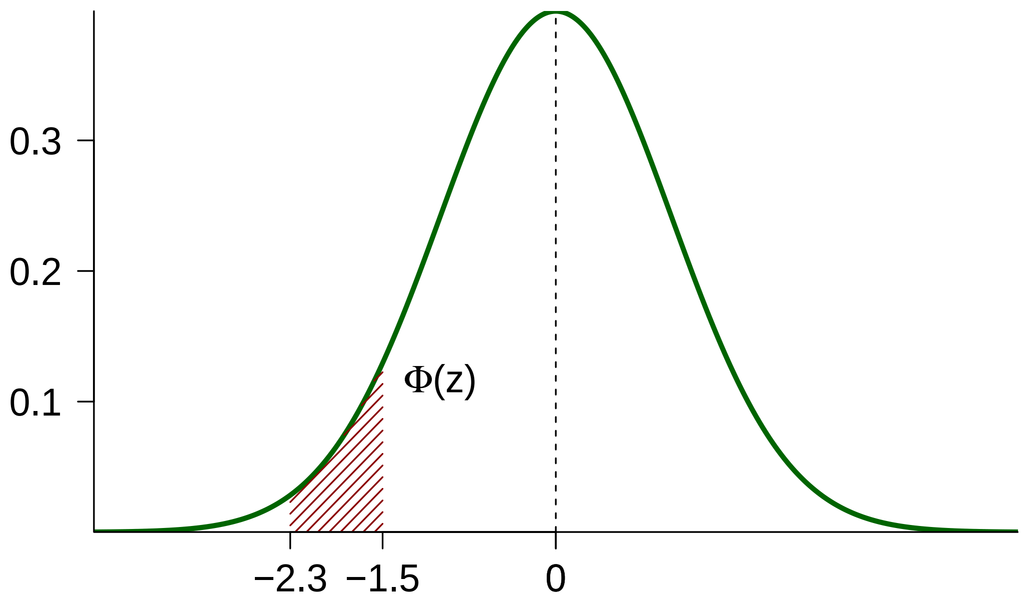
Pertanto cerchiamo sulle tavole $\Phi(1.96)$ incrociando la riga 1.9 con la colonna 0.06 e troviamo che la probabilità è 0.9750. Quindi:

$$P(Z \geq 1.96) = 1 - P(Z \leq 1.96) = 1 - \Phi(1.96) = 1 - 0.975 = 0.025.$$

La probabilità di osservare un valore superiore o uguale a 1.96 nella normale standard è di 0.025 (cioè del 2.5%).

Es. h)

Mediante la funzione di ripartizione della normale standard è possibile anche calcolare qual è la probabilità che si osservino valori compresi tra due estremi. Basta effettuare opportune differenze tra aree (cioè tra integrali). Ad esempio la probabilità $P(-2.3 \leq Z \leq -1.5)$ corrisponde alla seguente area:



Ragioniamo in termini di differenze di aree. In definitiva l'area cercata è la **differenza tra l'area fino a z=-1.5 e l'area fino a z=-2.3**. In termini di funzione di ripartizione della normale standard tali aree sono rispettivamente:

$$P(Z \leq -1.5) = 1 - P(Z \leq 1.5) = 1 - \Phi(1.5)$$

e

$$P(Z \leq -2.3) = 1 - P(Z \leq 2.3) = 1 - \Phi(2.3)$$

combinando le due:

$$P(-2.3 \leq Z \leq -1.5) = 1 - \Phi(1.5) - [1 - \Phi(2.3)] = 1 - \Phi(1.5) - 1 + \Phi(2.3) = \Phi(2.3) - \Phi(1.5)$$

Andiamo a cercare tali integrali sulle tavole: $\Phi(2.3) = 0.9893$ e $\Phi(1.5) = 0.9332$. Quindi:

$$P(-2.3 \leq Z \leq -1.5) = 0.9893 - 0.9332 = 0.0561$$

Nella normale standard, la probabilità di osservare dei valori di z compresi tra -2.3 e -1.5 è di 0.0561.

Esercizio 4

Il tempo necessario per completare questa esercitazione segue una distribuzione normale di media 100 minuti e deviazione standard di 20 minuti.

a) Calcolare con che probabilità gli studenti completeranno tutti gli esercizi entro 2 ore

b) Calcolare con che probabilità gli studenti completeranno tutti gli esercizi esattamente in 2 ore

c) Quanto tempo è necessario affinché il 95% degli studenti completino l'esercitazione?

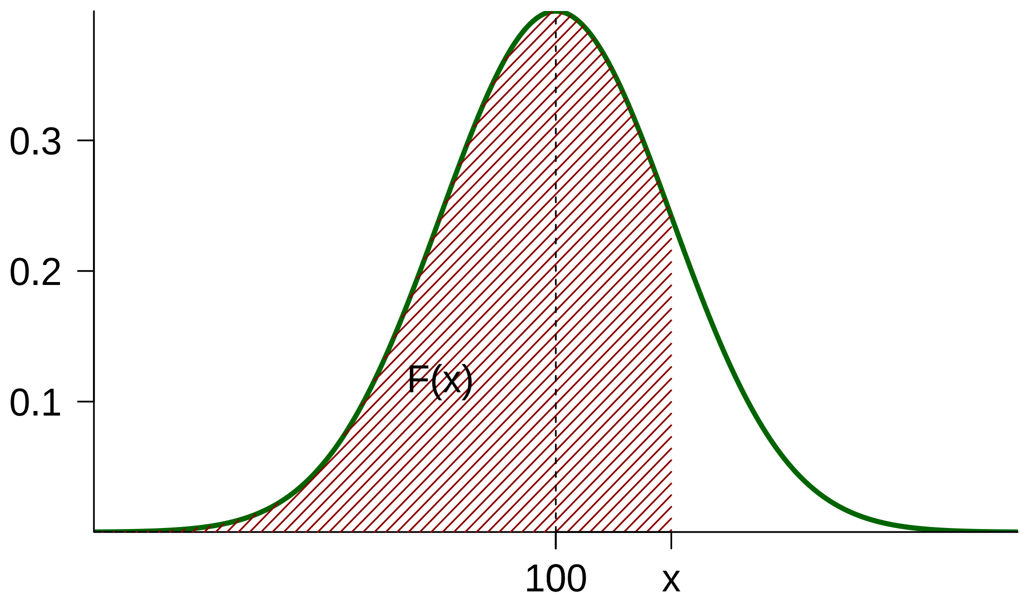
Es. a)
Dai dati forniti il tempo per completare questa esercitazione è una variabile aleatoria X che si distribuisce come un normale, cioè $X \sim Nor(100, 20)$.
Ci chiediamo qual è la probabilità che gli studenti completino l'esercitazione in 2 ore ossia in un tempo minore o uguale ai 120 minuti esprimendo il tempo nell'unità di misura di X.
Come è noto non sappiamo calcolare la $F(x)$ per qualsiasi normale con parametri arbitrari ma solo la $F(x)$ della particolare normale con media 0 e deviazione standard 1 ossia la $\Phi()$ della distribuzione $Z \sim Nor(0, 1)$. Pertanto quando vogliamo delle probabilità di una normale generica dobbiamo sempre ridurci al caso della Z, ossia trasformare la nostra X che ha $\mu = 100$ e $\sigma = 20$ in una variabile aleatoria con $\mu = 0$ e $\sigma = 1$.
In particolare dobbiamo usare la trasformazione detta **standardizzazione**.
In simboli noi vogliamo la seguente probabilità in termini di X, $P(X < 120)$ ossia $F(120)$.
Per calcolare $F(120)$ occorre standardizzare 120. Così troveremo l'area sotto la curva corrispondente ad una normale standard (ricordatevi che le aree non cambiano anche se operiamo tale trasformazione). Per cui applicando la standardizzazione ad entrambi i termini in $P()$:

$$P(X < 120) = P(\frac{X-\mu}{\sigma} < \frac{120-\mu}{\sigma}) = P(\frac{X-100}{20} < \frac{120-100}{20}) =$$

standardizzare in generale la variabile X (primo termine della disuguaglianza in $P()$) equivale a trasformare in generale la variabile X nella Z:

$$= P(Z < \frac{120-100}{20}) = P(Z < 1) = \Phi(1) = 0.8413$$

Cioè vi è circa l'84% di probabilità che gli studenti completino questo esercizitazione entro le 2 ore.



Es. b)
Per le variabili aleatorie continue si ha $P(X = x) = 0$ pertanto la probabilità che gli studenti completeranno tutti gli esercizi esattamente in 2 ore è nulla.
Es. c)
Per calcolare il tempo (quindi un valore x della variabile X) necessario affinchè il 95% degli studenti completino l'esercitazione dobbiamo prima trovare il quantile q (o percentile) dalla distribuzione standardizzata che sottende ad un'area del 95% (cioè quel valore z_q tale per cui $\Phi(z_q) = 0.95$) e poi applicare la trasformazione inversa alla srandardizzazione per passare dal valore z a quello x della variabile originaria (non trasformata).
Dalle tavole cercando un valore nele celle approssimativamente pari a 0.95 (e risalendo poi ai valori in riga e in colonna), si ha che:

$$P(Z < z_q) = \Phi(z_q) = 0.95 \Rightarrow z_q \approx 1.645$$

Successivamente per trovare il tempo x utilizziamo la trasformazione inversa della standardizzazione, cioè aggiungiamo μ e moltiplichiamo per σ sia Z che il valore individuato $z_q = 1.645$:

$$P(Z < z_q) = 0.95 \Rightarrow \Phi(Z * 20 + 100 < z_q * 20 + 100) = 0.95 \Rightarrow \Phi(X < x_q) = 0.95$$

quindi x sarà:

$$x_q = z_q * 20 + 100 = 1.645 * 20 + 100 = 132.9$$

Sono necessari circa 133 minuti affinché il 95% degli studenti completi l'esercitazione.

Esercizio 5

In una popolazione di studenti di una determinata scuola è noto che i loro punteggi ai test per il calcolo del quoziente di intelligenza (QI) seguono una distribuzione normale avente media 95 e varianza 225. Calcolare la probabilità di osservare uno studente con:

a) QI maggiore di 95

b) QI minore di 90

c) QI maggiore di 90

d) QI compreso tra 85 e 110

e) QI minore di 92 oppure maggiore di 100

f) Qual è il punteggio QI ottenuto dal 50% degli studenti di quella scuola?

g) Qual è il punteggio QI ottenuto dal 90% degli studenti di quella scuola?

Suggerimento: come prima cosa individuate i parametri di X, ossia scrivete la seguente espressione $X \sim Nor(\mu, \sigma)$ facendo attenzione a quali dati vi sono stati forniti!