PROBABILITÀ – SCHEDA N. 7
LA VARIABILE ALEATORIA NORMALE

Molti fenomeni naturali come le altezze, i pesi, le misure dei pezzi prodotti da una fabbrica secondo determinate specifiche, gli errori casuali nelle misurazioni fisiche, ... sono modellabili (almeno in prima approssimazione) con una famiglia di variabili aleatorie, dette *normali (o gaussiane)*, la cui densità di probabilità ha una forma cosiddetta "a campana" (simmetrica rispetto alla media, che assume molti valori vicino alla media, ...) come indicato nella Figura 1.

Una variabile aleatoria **normale** X assume valori su tutta la retta dei numeri reali. Se ha valore atteso μ e varianza σ^2 la sua densità è

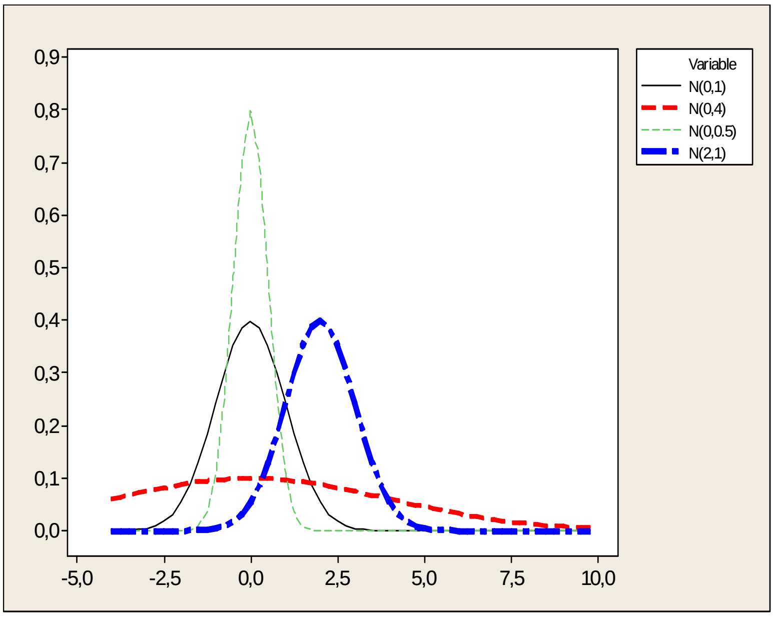
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

e si indica con

$$X \sim N(\mu, \sigma^2).$$

Quando il valore atteso è uguale a 0 e la varianza 1 si parla di **normale standardizzata**.

Qui di seguito sono riportati alcuni grafici delle densità normali al variare del valore atteso e della varianza.



Si può osservare che:

- il grafico è simmetrico rispetto al valore atteso;
- densità con uguale varianza, ma media diversa, hanno la stessa forma ma sono traslate;
- la varianza fa cambiare la forma del grafico; più la varianza cresce più i valori sono dispersi rispetto alla media e il grafico "si appiattisce"; viceversa se la varianza è piccola, i valori sono più addensati intorno alla media e il grafico "è appuntito".

Figura 1. Densità di $X \sim N(\mu, \sigma^2)$

La famiglia delle variabili aleatorie normali gode di buone proprietà matematiche, molto utili nelle applicazioni. Ne vediamo alcune in questa scheda e altre nella prossima.

1. Una variabile aleatoria Y ottenuta con una *trasformazione lineare* di una v.a. X con distribuzione $N(\mu, \sigma^2)$:
$$Y = aX + b$$
è ancora una normale, con valore atteso e varianza opportunamente trasformate:
$$Y \sim N(a\mu + b, a^2\sigma^2)$$
In particolare se $X \sim N(0,1)$ allora $Y \sim N(b, a^2\sigma^2)$.
[Quale altra variabile aleatoria ha la stessa proprietà?]

2. La *somma* di variabili aleatorie con distribuzione normale è una variabile aleatoria con distribuzione normale (senza ipotesi di indipendenza).

3. Se due variabili normali sono non correlate, allora sono indipendenti. L'indipendenza è una condizione più forte della non correlazione, infatti l'indipendenza riguarda la distribuzione delle variabili aleatorie, mentre la non correlazione coinvolge solo i loro valori attesi.

Calcolo di $P(a < X < b)$ **.**

La prima proprietà è fondamentale per il calcolo della probabilità che la variabile appartenga a un intervallo. Infatti abbiamo già visto che le probabilità si calcolano come:

$$P(a < X < b) = \int_a^b f_X(x)\,dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \,.$$

Poiché non esiste una primitiva della funzione di densità $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, NON si possono calcolare esplicitamente gli integrali di f_X su un generico intervallo (a,b).

Per ovviare a questo problema, esistono tavole (riportate in fondo a questa scheda) che forniscono la funzione di distribuzione cumulata per la normale standardizzata $N(0,1)$ e da queste si ricavano le informazioni per tutte le altre variabili aleatorie normali.

Infatti se X è una variabile aleatoria normale con media μ e varianza σ^2, la variabile Z definita da

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

è ancora una variabile aleatoria normale (essendo una trasformazione lineare di X) con media 0 e varianza 1: $Z \sim N(0,1)$. Infatti:

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = 0$$

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X-\mu) = \frac{1}{\sigma^2}Var(X) = 1$$

La trasformazione di X in Z permette di calcolare $P(x_1 < X < x_2)$ usando $P(z_1 < Z < z_2)$ e quindi di utilizzare le tavole.

ESEMPIO 1. Consideriamo $X \sim N(30,4)$ e calcoliamo $P(28.5 < X < 34)$.

In questo caso $Z = \frac{X-30}{2}$ con $Z \sim N(0,1)$.

Per i valori di X uguali a 28.5 e 34, calcoliamo i corrispondenti valori z di una variabile aleatoria Z:

- se x = 28, allora z = (28.5 - 30) / $\sqrt{4}$ = -0.75
- se x = 34, allora z = (34 - 30) / $\sqrt{4}$ = 2

Quindi:

$$P(28.5 < X < 34) = P\left(\frac{28.5-30}{\sqrt{4}} < \frac{X-30}{\sqrt{4}} < \frac{34-30}{\sqrt{4}}\right) = P(-0.75 < Z < 2) =$$

$$= P(Z < 2) - P(Z < -0.75) = F_Z(2) - F_Z(-0.75).$$

Uso delle tavole di una variabile aleatoria con distribuzione normale standardizzata

Le tavole riportate in Appendice relative alla funzione di distribuzione cumulata di una variabile casuale con distribuzione *N(0,1)* si leggono nel seguente modo: i numeri posti nella prima colonna corrispondono ai valori z di Z con una cifra decimale; la seconda cifra decimale si legge nella prima riga. All'incrocio tra riga e colonna si ha il valore della funzione di distribuzione cumulata.

Alcuni calcoli.

a) Per calcolare il valore della funzione di distribuzione cumulata in z = -0.75, cioè $P(Z \le -0.75)$, ovvero l'area della curva della densità a sinistra del valore -0.75, dobbiamo cercare nella prima colonna della tabella il valore -0.7 e poi nella prima riga il valore 5: nel punto di incrocio di questa riga e questa colonna troviamo 0.2266; quindi: $P(Z \le -0.75)$ = 0.2266.

b) Per calcolare $P(Z \le 2.00)$ dobbiamo cercare nella prima colonna della tabella il valore 2.0 e poi nella prima riga il valore 0: nel punto di incrocio di questa riga e questa colonna troviamo 0.9772; quindi: $P(Z \le 2.00)$ = 0.9772.

Il risultato dell'Esempio 1 è quindi: $P(28.5 < X < 34) = F_Z(2) - F_Z(-0.75) =$ 0.9772 – 0.2266 = 0.7506.

c) Per calcolare $P(Z > 0.43)$ possiamo cercare $P(Z < 0.43)$ e quindi $P(Z > 0.43)$ = 1 – $P(Z < 0.43)$. Si trova: $P(Z > 0.43)$ = 0.6664 quindi $P(Z > 0.43)$ = 0.3336.
Esiste un'altra strada che sfrutta la simmetria della funzione di densità.

$$P(Z > 0.43) = P(Z < -0.43) = 0.3336$$

N.B. Le disuguaglianze si possono porre indifferentemente larghe ($\le$, $\ge$) o strette (<, >).

Le tavole possono essere utilizzate anche per calcolare i valori dei quantili di X. La trasformazione inversa da Z a X è: $X = \sigma Z + \mu$.

ESEMPIO 2. Consideriamo sempre $X \sim N(30, 4)$ e calcoliamo il primo e il terzo quartile di X, cioè il valori x_{25} e x_{75} tali che:

$$P(X < x_{25}) = 0.25 \text{ e } P(X < x_{75}) = 0.75.$$

La trasformazione inversa da Z a X è:

$$X = 2\,Z + 30.$$

Sulle tavole si trova $0.25 = P(Z < -0.68)$. Quindi

$$0.25 = P(Z < -0.68) = P(2 \times Z + 30 < -2 \times 0.68 + 30) = P(X < -2 \times 0.68 + 30) = P(X < 28.64)$$

Analogamente:

$$0.75 = P(Z < 0.68) = P(2 \times Z + 30 < 2 \times 0.68 + 30) = P(X < 2 \times 0.68 + 30) = P(X < 31.36).$$

Quindi: x_{25} = 28.64 e x_{75} =31.36.

ESEMPIO 3. I diametri delle aste prodotte da una certa macchina sono modellabili con una variabile aleatoria X con distribuzione normale con media 10 cm e deviazione standard 0.1 cm. Se per un dato uso, un'asta deve avere un diametro compreso fra 9.9 e 10.2 cm, qual è la probabilità che un'asta prodotta da questa macchina sia in grado di soddisfare tale requisito?
Dobbiamo quindi calcolare $P(9.9 < X < 10.2)$.

$$P(9.2 < X < 10.1) = P\left(\frac{9.9-10}{0.1} < \frac{X-10}{0.1} < \frac{10.2-10}{0.1}\right) = P\left(\frac{9.9-10}{0.1} < Z < \frac{10.2-10}{0.1}\right)$$
$$= P(-1 < Z < 2) = P(Z < 2) - P(Z < -1) = 0.9772 - 0.1587 = 0.8185.$$

ESEMPIO 4: Un tipo di conserve è inscatolato in lattine di peso medio 400 grammi. Il confezionamento viene effettuato da una macchina che è tarata a 405 grammi (altrimenti metà delle confezioni avrebbero peso inferiore a quello dichiarato); supponiamo che il peso delle conserve sia modellabile con una normale di deviazione standard fissa 3 grammi. In un lotto di 10000 lattine quante si prevede che siano di peso inferiore a 400 grammi? Indichiamo con X il peso delle conserve in una lattina: X è una variabile normale $N(405, 9)$. Dobbiamo calcolare P(X < 400) utilizzando le tavole della normale standardizzata. Abbiamo che

$$P(X < 400) = P\left(\frac{X-405}{3} < \frac{400-405}{3}\right) = P(Z < -1.67) = 0.0475\,.$$

Se indichiamo con Y il numero di lattine con peso inferiore a 400 grammi in un lotto di 10000 lattine, Y è una variabile binomiale B(10000, 0.0475). Allora il numero medio di quelle con peso inferiore a quello dichiarato è 475 (=10000x0.0475)

Supponiamo ora di voler cambiare la taratura della macchina in modo tale che la probabilità che una lattina abbia peso inferiore a 400 grammi sia minore del 1% (mantenendo lo stesso scarto). In questo caso X è una normale $N(\mu, 9)$. Come prima abbiamo che

$$\mathrm{P(X < 400)} = P\left(\frac{X-\mu}{3} < \frac{400-\mu}{3}\right) = P\left(Z < \frac{400-\mu}{3}\right)$$

Se imponiamo che questa quantità sia minore di 0.01, abbiamo (sempre leggendo dalle tavole) che $\frac{400-\mu}{3} < -2.33$, da cui si ricava che μ > 406.99. Quindi la macchina deve essere tarata a 407 grammi.

ESERCIZI

ESERCIZIO 1. Una popolazione è costituita da tre sottopopolazioni che indicheremo con P_1, P_2 e P_3. La proporzione delle tre popolazioni è rispettivamente 0.15, 0.25 e 0.60. In ciascuna di queste popolazioni viene rilevata una caratteristica qualitativa A. Si trova che il 20% di P_1, l'8% di P_2 e il 5% di P_3 presentano la caratteristica A.

a) Qual è la frequenza degli individui che nella popolazione totale presentano la caratteristica A?
   Dalla popolazione si estrae un campione casuale di numerosità 100.
b) In media quanti individui con la caratteristica A si troveranno in un tale campione?
c Qual è la varianza del numero di individui con la caratteristica A in tale campione?
d) Qual è la probabilità di trovare meno di 5 individui con la caratteristica A (approssimare eventualmente la distribuzione)?

ESERCIZIO 2. Un giardiniere decide di piantare delle petunie in un'aiuola; compra un pacco di semi misti di petunie sulla cui etichetta è scritto che 1/5 dei semi produrranno petunie rosa e 4/5 produrranno petunie rosse e bianche. Assumiamo che i semi siano stati scelti a caso da una popolazione mista secondo le proporzioni indicate e che il modello binomiale sia applicabile. Se germinano 200 semi, qual è la probabilità di avere almeno 30 ma non più di 50 petunie rosa?

ESERCIZIO 3. Il signor Rossi si deve presentare al luogo di lavoro entro le ore 9. Egli esce di casa alle 8 esatte e si reca camminando alla fermata dell'autobus che, sempre puntuale!, passa alle 8.15; il tempo T1 che egli impiega a compiere questo tragitto può essere modellato con una variabile aleatoria normale di media 10 min e deviazione standard 2 min. Il tempo T2 che l'autobus impiega per portare il signor Rossi sul posto di lavoro si può guardare come una variabile aleatoria normale di media 38 min e scarto 5 min.

a) Qual è la probabilità che il signor Rossi perda l'autobus?

b) Se il signor Rossi arriva in tempo per prendere l'autobus, qual è la probabilità che arrivi puntuale sul posto di lavoro?

c) Nel momento in cui il signor Rossi esce di casa, qual è la probabilità che arrivi puntuale sul posto di lavoro?

ESERCIZIO 4. Consideriamo variabile aleatoria X con $X \sim N(\mu, \sigma^2)$. Calcolare le seguenti probabilità legate ai valori di μ e σ:

a) $P(\mu - \sigma < X < \mu + \sigma)$ b) $P(\mu - 2\sigma < X < \mu + 2\sigma)$ c) $P(\mu - 3\sigma < X < \mu + 3\sigma)$

ESERCIZIO 5. Consideriamo variabile aleatoria Z con $Z \sim N(0,1)$. Cercare sulle tavole quei valori (approssimati) z di Z che comprendono il 50%, il 95% e il 99% dei dati centrali, cioè

a) $P(-z < Z < z) =$ 0.50 b) $P(-z < Z < z) = 0.95$ c) $P(-z < Z < z) = 0.99$

E1. Esperienza. La variabile aleatoria normale

a) Tabulare la funzione densità normale $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ nell'intervallo $I = (\mu - 3\sigma\ ;\ \mu + 3\sigma\ )$, con passo *p=1/100* e assegnando i valori di μ e di σ in input.

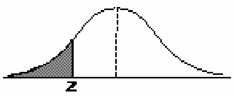
b) Tracciare il grafico di $f_X(x)$ e osservare come si modifica al variare dei valori di μ e di σ

c) Scegliendo μ =0 e σ =1 ottenere la densità normale standard $\phi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$ tabularla e tracciarne il grafico

d) Approssimare l'integrale di $\phi(z)$ nell'intervallo *I* con le somme integrali per difetto e per eccesso del metodo di Cauchy-Riemann e utilizzando il passo *p* usato per la tabulazione

Tavole della funzione di distribuzione cumulata di Z, con $Z \sim N(0,1)$

$$P(Z \le z) = F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx$$



z	0	1	2	3	4	5	6	7	8	9
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
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