CHAPTER 8

Normal distribution

8.1. Standard and general normal distributions

Definition (Standard normal distribution)

A continuous random variable is a *standard normal* (written $\mathcal{N}(0,1)$) if it has density

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

A synonym for normal is Gaussian. The first thing to do is to show that this is a (probability) density.

Theorem

$f_Z(x)$ is a valid PDF, that is, it is a nonnegative function such that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1.$$

Suppose $Z \sim \mathcal{N}(0,1)$. Then

$$\mathbb{E}Z = 0,$$
$$\operatorname{Var} Z = 1.$$

Proof. Let $I = \int_0^\infty e^{-x^2/2} dx$. Then

$$I^2 = \int_0^\infty \int_0^\infty e^{-x^2/2} e^{-y^2/2} dx\, dy.$$

Changing to polar coordinates,

$$I^2 = \int_0^{\pi/2} \int_0^\infty r e^{-r^2/2} dr = \pi/2.$$

So $I = \sqrt{\pi/2}$, hence $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ as it should.

Note

$$\int_{-\infty}^{\infty} x e^{-x^2/2} dx = 0$$

by symmetry, so $\mathbb{E}Z = 0$. For the variance of Z, we use integration by parts as follows.

$$\mathbb{E}Z^2 = \frac{1}{\sqrt{2\pi}} \int x^2 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int x \cdot x e^{-x^2/2} dx.$$

The integral is equal to

$$-xe^{-x^2/2}\Big]_{-\infty}^{\infty}+\int e^{-x^2/2}dx=\sqrt{2\pi}.$$

Therefore $\operatorname{Var} Z=\mathbb{E}Z^2=1$. □

Note that these integrals are improper, so our arguments are somewhat informal, but they are easy to make rigorous.

Definition (General normal distribution)

We say X is a $\mathcal{N}(\mu,\sigma^2)$ if $X=\sigma Z+\mu$, where $Z\sim\mathcal{N}(0,1)$.

Proposition 8.1 (General normal density)

A random variable with a general normal distribution $\mathcal{N}(\mu,\sigma^2)$ is a continuous random variable with density

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

Proof. Suppose $X\sim\mathcal{N}(\mu,\sigma^2)$. We see that if $\sigma>0$, then

$$\begin{aligned}F_X(x)&=\mathbb{P}(X\leqslant x)=\mathbb{P}(\mu+\sigma Z\leqslant x)\\&=\mathbb{P}\left(Z\leqslant\frac{x-\mu}{\sigma}\right)=F_Z\left(\frac{x-\mu}{\sigma}\right),\end{aligned}$$

where $Z\sim\mathcal{N}(0,1)$. A similar calculation holds if $\sigma<0$.

Then by the chain rule X has density

$$f_X(x)=F_X'(x)=F_Z'\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sigma}f_Z\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

□

Proposition 8.2 (Properties of general normal variables)

For any $X\sim\mathcal{N}(\mu,\sigma)$

$$\begin{aligned}\mathbb{E}X&=\mu,\\\operatorname{Var}X&=\sigma^2.\end{aligned}$$

For any $a,b\in\mathbb{R}$ the random variable $aX+b$ is a normal variable.

Proof. By definition of X we have $X=\sigma Z+\mu$, where $Z\sim\mathcal{N}(0,1)$, and so by linearity of the expectation we have $\mathbb{E}X=\mu+\mathbb{E}Z=\mu$. Recall that by Exercise 7.1 we have $\operatorname{Var}X=\sigma^2\operatorname{Var}Z$, so

$$\operatorname{Var}X=\sigma^2.$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$, then $Y = a(\mu + \sigma Z) + b = (a\mu + b) + (a\sigma)Z$, or Y is $\mathcal{N}(a\mu + b, a^2\sigma^2)$. In particular, if $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z = (X - \mu)/\sigma$, then Z is $\mathcal{N}(0, 1)$. $\square$

The distribution function of a standard random variable $\mathcal{N}(0, 1)$ is often denoted as $\Phi(x)$, so that

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

Tables of $\Phi(x)$ are often given only for $x > 0$. One can use the symmetry of the density function to see that

$$\Phi(-x) = 1 - \Phi(x),$$

this follows from

$$\begin{aligned}
\Phi(-x) = \mathbb{P}(Z \leqslant -x) &= \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
&= \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \mathbb{P}(Z \geqslant x) \\
&= 1 - \mathbb{P}(Z < x) = 1 - \Phi(x).
\end{aligned}$$

Table 1. Cumulative distribution function for the standard normal variable

x	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.9998	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.9999	0.9999	0.9999	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

Example 8.1. Find $\mathbb{P}(1 \leqslant X \leqslant 4)$ if X is $\mathcal{N}(2, 25)$.

Solution: Write $X = 2 + 5Z$, so

$$\begin{aligned}\mathbb{P}(1 \leqslant X \leqslant 4) &= \mathbb{P}(1 \leqslant 2 + 5Z \leqslant 4) \\ &= \mathbb{P}(-1 \leqslant 5Z \leqslant 2) = \mathbb{P}(-0.2 \leqslant Z \leqslant 0.4) \\ &= \mathbb{P}(Z \leqslant 0.4) - \mathbb{P}(Z \leqslant -0.2) \\ &= \Phi(0.4) - \Phi(-0.2) = 0.6554 - (1 - \Phi(0.2)) \\ &= 0.6554 - (1 - 0.5793)\,.\end{aligned}$$

Example 8.2. Find c such that $\mathbb{P}(|Z| \geqslant c) = 0.05$.

Solution: By symmetry we want to find c such that $\mathbb{P}(Z \geqslant c) = 0.025$ or $\Phi(c) = \mathbb{P}(Z \leqslant c) = 0.975$. From the table (Table 1) we see $c = 1.96 \approx 2$. This is the origin of the idea that the 95% significance level is ± 2 standard deviations from the mean.

Proposition 8.3

For a standard normal random variable Z we have the following bound. For $x > 0$

$$\mathbb{P}(Z \geqslant x) = 1 - \Phi(x) \leqslant \frac{1}{\sqrt{2\pi}\ x} e^{-x^2/2}.$$

Proof. If $y \geqslant x > 0$, then $y/x \geqslant 1$, and therefore

$$\begin{aligned}\mathbb{P}(Z \geqslant x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\, dy \\ &\leqslant \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{y}{x} e^{-y^2/2}\, dy = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.\end{aligned}$$

□

This is a good estimate when x is large. In particular, for x large,

$$\mathbb{P}(Z \geqslant x) = 1 - \Phi(x) \leqslant e^{-x^2/2}.$$

8.2. Further examples and applications

Example 8.3. Suppose X is normal with mean 6. If $\mathbb{P}(X > 16) = 0.0228$, then what is the standard deviation of X?

Solution: recall that we saw in Proposition 8.2 that $\frac{X-\mu}{\sigma} = Z \sim \mathcal{N}(0,1)$ and then

$$\begin{aligned}
\mathbb{P}(X > 16) = 0.0228 &\iff \mathbb{P}\left(\frac{X-6}{\sigma} > \frac{16-6}{\sigma}\right) = 0.0228 \\
&\iff \mathbb{P}\left(Z > \frac{10}{\sigma}\right) = 0.0228 \\
&\iff 1 - \mathbb{P}\left(Z \leqslant \frac{10}{\sigma}\right) = 0.0228 \\
&\iff 1 - \Phi\left(\frac{10}{\sigma}\right) = 0.0228 \\
&\iff \Phi\left(\frac{10}{\sigma}\right) = 0.9772.
\end{aligned}$$

Using the standard normal table we see that $\Phi(2) = 0.9772$, thus we have that

$$2 = \frac{10}{\sigma}$$

and hence $\sigma = 5$.

8.3. Exercises

Exercise 8.1. Suppose X is a normally distributed random variable with $\mu = 10$ and $\sigma^2 = 36$. Find (a) $\mathbb{P}(X > 5)$, (b) $\mathbb{P}(4 < X < 16)$, (c) $\mathbb{P}(X < 8)$.

Exercise 8.2. The height of maple trees at age 10 are estimated to be normally distributed with mean 200 cm and variance 64 cm. What is the probability a maple tree at age 10 grows more than 210cm?

Exercise 8.3. The peak temperature T, in degrees Fahrenheit, on a July day in Antarctica is a Normal random variable with a variance of 225. With probability .5, the temperature T exceeds 10 degrees.

(a) What is $\mathbb{P}(T > 32)$, the probability the temperature is above freezing?
(b) What is $\mathbb{P}(T < 0)$?

Exercise 8.4. The salaries of UConn professors is approximately normally distributed. Suppose you know that 33 percent of professors earn less than \$80,000. Also 33 percent earn more than \$120,000.

(a) What is the probability that a UConn professor makes more than \$100,000?
(b) What is the probability that a UConn professor makes between \$70,000 and \$80,000?

Exercise 8.5. Suppose X is a normal random variable with mean 5. If $\mathbb{P}(X > 0) = 0.8888$, approximately what is $\operatorname{Var}(X)$?

Exercise 8.6. The shoe size of a UConn basketball player is normally distributed with mean 12 inches and variance 4 inches. Ten percent of all UConn basketball players have a shoe size greater than c inches. Find the value of c.

Exercise 8.7. The length of the forearm of a UConn football player is normally distributed with mean 12 inches. If ten percent of the football team players have a forearm whose length is greater than 12.5 inches, find out the approximate standard deviation of the forearm length of a UConn football player.

Exercise 8.8. Companies C and A earn each an annual profit that is normally distributed with the same positive mean μ. The standard deviation of C's annual profit is one third of its mean. In a certain year, the probability that A makes a loss (i.e. a negative profit) is 0.8 times the probability that C does. Assuming that A's annual profit has a standard deviation of 10, compute (approximately) the standard deviation of C's annual profit.

Exercise 8.9. Let $Z \sim \mathcal{N}(0, 1)$, that is, a standard normal random variable. Find probability density for $X = Z^2$. *Hint:* first find the (cumulative) distribution function $F_X(x) = \mathbb{P}(X \leqslant x)$ in terms of $\Phi(x) = F_Z(x)$. Then use the fact that the probability density function can be found by $f_X(x) = F'_X(x)$, and use the known density function for Z.

8.4. Selected solutions

Solution to Exercise 8.1(A):

$$\begin{aligned}\mathbb{P}(X>5)&=\mathbb{P}\left(Z>\frac{5-10}{6}\right)=\mathbb{P}\left(Z>-\frac{5}{6}\right)\\&=\mathbb{P}\left(Z<\frac{5}{6}\right)=\Phi\left(\frac{5}{6}\right)\approx\Phi(0.83)\approx 0.797\end{aligned}$$

Solution to Exercise 8.1(B): $2\Phi(1)-1=0.6827$

Solution to Exercise 8.1(C): $1-\Phi(0.3333)=0.3695$

Solution to Exercise 8.2: We have $\mu=200$ and $\sigma=\sqrt{64}=8$. Then

$$\begin{aligned}\mathbb{P}(X>210)&=\mathbb{P}\left(Z>\frac{210-200}{8}\right)=\mathbb{P}(Z>1.25)\\&=1-\Phi(1.25)=0.1056.\end{aligned}$$

Solution to Exercise 8.3(A): We have $\sigma=\sqrt{225}=15$. Since $\mathbb{P}(X>10)=0.5$ then we must have that $\mu=10$ since the PDF of the normal distribution is symmetric. Then

$$\begin{aligned}\mathbb{P}(T>32)&=\mathbb{P}\left(Z>\frac{32-10}{15}\right)\\&=1-\Phi(1.47)=0.0708.\end{aligned}$$

Solution to Exercise 8.3(B): We have $\mathbb{P}(T<0)=\Phi(-0.67)=1-\Phi(0.67)=0.2514$.

Solution to Exercise 8.4(A): First we need to figure out what μ and σ are. Note that

$$\begin{aligned}\mathbb{P}(X\leqslant 80,000)=0.33&\iff\mathbb{P}\left(Z<\frac{80,000-\mu}{\sigma}\right)=0.33\\&\iff\Phi\left(\frac{80,000-\mu}{\sigma}\right)=0.33\end{aligned}$$

and since $\Phi(0.44)=0.67$ then $\Phi(-0.44)=0.33$. Then we must have

$$\frac{80,000-\mu}{\sigma}=-0.44.$$

Similarly, since

$$\begin{aligned}\mathbb{P}(X>120,000)=0.33&\iff 1-\mathbb{P}(X\leqslant 120,000)=0.33\\&\iff 1-\Phi\left(\frac{120,000-\mu}{\sigma}\right)=0.33\\&\iff\Phi\left(\frac{120,000-\mu}{\sigma}\right)=0.67\end{aligned}$$

Now again since $\Phi(0.44)=0.67$ then

$$\frac{120,000-\mu}{\sigma}=0.44.$$

Solving the equations

$$\frac{80,000-\mu}{\sigma}=-0.44 \text{ and } \frac{120,000-\mu}{\sigma}=0.44,$$

as a system for μ and σ we have that

$$\mu=100,000 \text{ and } \sigma\approx 45,454.5.$$

Then

$$\mathbb{P}(X>100,000)=0.5.$$

Solution to Exercise 8.4(B): We have

$$\mathbb{P}(70,000<X<80,000)\approx 0.0753.$$

Solution to Exercise 8.5: Since $\mathbb{P}(X>0)=.8888$, then

$$\begin{aligned}
\mathbb{P}(X>0)=0.8888 &\iff \mathbb{P}\left(Z>\frac{0-5}{\sigma}\right)=0.8888\\
&\iff 1-\mathbb{P}\left(Z\leqslant-\frac{5}{\sigma}\right)=0.8888\\
&\iff 1-\Phi\left(-\frac{5}{\sigma}\right)=0.8888\\
&\iff 1-\left(1-\Phi\left(\frac{5}{\sigma}\right)\right)=0.8888\\
&\iff \Phi\left(\frac{5}{\sigma}\right)=0.8888.
\end{aligned}$$

Using the table we see that $\Phi(1.22)=0.8888$, thus we must have that

$$\frac{5}{\sigma}=1.22$$

and solving this gets us $\sigma=4.098$, hence $\sigma^2\approx 16.8$.

Solution to Exercise 8.6: Note that

$$\begin{aligned}
\mathbb{P}(X>c)=0.10 &\iff \mathbb{P}\left(Z>\frac{c-12}{2}\right)=0.10\\
&\iff 1-\mathbb{P}\left(Z\leqslant\frac{c-12}{2}\right)=0.10\\
&\iff \mathbb{P}\left(Z\leqslant\frac{c-12}{2}\right)=0.9\\
&\iff \Phi\left(\frac{c-12}{2}\right)=0.9
\end{aligned}$$

Using the table we see that $\Phi(1.28)=0.90$, thus we must have that

$$\frac{c-12}{2}=1.28$$

and solving this gets us $c=14.56$.

Solution to Exercise 8.7: Let X denote the forearm length of a UConn football player and let σ denote its standard deviation. From the problem we know that

$$\mathbb{P}(X > 12.5) = \mathbb{P}\left(\frac{X-12}{\sigma} > \frac{0.5}{\sigma}\right) = 1 - \Phi\left(\frac{0.5}{\sigma}\right) = 0.1.$$

From the table we get $\frac{0.5}{\sigma} \approx 1.29$ hence $\sigma \approx 0.39$.

Solution to Exercise 8.8: Let A and C denote the respective annual profits, and μ their mean. Form the problem we know $\mathbb{P}(A < 0) = 0.8\mathbb{P}(C < 0)$ and $\sigma_A = \mu/3$. Since they are normal distributed, $\Phi\left(\frac{-\mu}{10}\right) = 0.8\Phi(-3)$ which implies

$$\Phi\left(\frac{\mu}{10}\right) = 0.2 + 0.8\Phi(3) \approx 0.998.$$

From the table we thus get $\mu/10 \approx 2.88$ and hence the standard deviation of C is $\mu/3 \approx 9.6$.

Solution to Exercise 8.9: see Example 10.2.