Normal probability density functions.

Recall:

(1) An rv X whose range is some interval (c,d) of real numbers is a continuous rv.

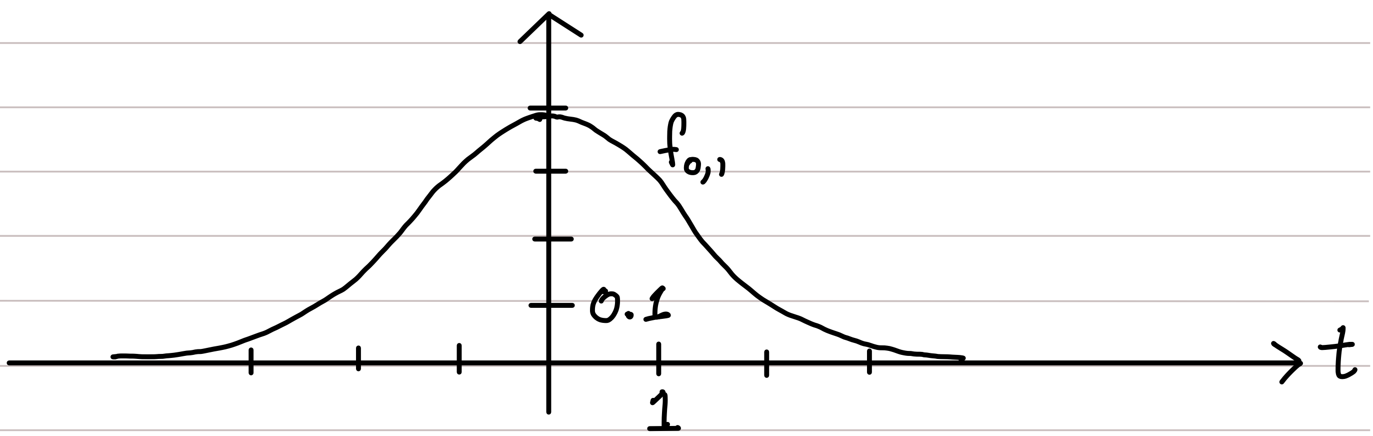
(2) The probability density function, or pdf, for such an X is a function f on (c,d) satisfying

$$P(a < X < b) = \int_a^b f(x)\,dx.$$

(A) The standard normal pdf.

Let $f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Then $f_{0,1}$ is a pdf (see HW10), with mean $\mu = 0$ and std dev $\sigma = 1$, and domain $(-\infty, \infty)$.



If X is a population or an rv with pdf $f_{0,1}$, we say "X is standard normal" or "X is $N(0,1)$."

Then of course, for $-\infty < a \le b < \infty$,

$$P(a < X < b) = \int_a^b f_{0,1}(x)\,dx = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,dx.$$

Note:

$f_{0,1}$ has no simple antiderivative, so in general, to compute standard normal probabilities, we need Riemann sums.

For example, it's known that, if $X = N(0,1)$, then

(i) $P(-1 < X < 1) = 0.683$. That is: in a standard normal distribution, 68.3% of the data is within one std. dev. of the mean.

Similarly,

$$P(-2 < X < 2) = 0.955,$$
$$P(-3 < X < 3) = 0.997.$$

(ii) $P(-1.96 < X < 1.96) = 0.95$. In a standard normal distribution, 95% of the data is within 1.96 std devs. of the mean.

Similarly,

$$P(-2.33 < X < 2.33) = 0.98,$$
$$P(-2.58 < X < 2.58) = 0.99.$$

(B) <u>Other normal pdf's.</u> If

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

then one checks that $f_{\mu,\sigma}$ is a pdf, with mean μ and std dev σ. (And $f_{\mu,\sigma}$ is normal — it has the same basic shape as $f_{0,1}$.)

If Z is an rv with pdf $f_{\mu,\sigma}$, we say "Z is $N(\mu,\sigma)$".

Useful idea: to know $N(\mu,\sigma)$ probabilities, it's enough to know $N(0,1)$ probabilities.

Here's why:

<u>NISNID (normal is standard normal in disguise) fact:</u>

If Z is $N(\mu,\sigma)$, then

$$X = \frac{Z-\mu}{\sigma} \quad \text{is } N(0,1).$$

[Proof omitted.]

Here's how we use this fact:

<u>Examples.</u>

(1) Suppose Z is $N(2, 0.5)$. Find $P(1.02 < Z < 2.98)$.

<u>Solution.</u>

"standardize" Z (subtract μ, then divide by σ)

$$P(1.02 < Z < 2.98) = P\left(\frac{1.02-2}{0.5} < \frac{Z-2}{0.5} < \frac{2.98-2}{0.5}\right)$$

$$= P\left(-1.96 < \frac{Z-2}{0.5} < 1.96\right) = 0.95 = 95\%.$$

by NISNID fact and by A(ii) above

(2) Suppose Y is $N(-1,3)$. Find $P(-4<Y<2)$.

Solution.

$$P(-4<Y<2) = P\left(\frac{-4-(-1)}{3} < \frac{Y-(-1)}{3} < \frac{2-(-1)}{3}\right)$$

$$= P\left(-1 < \frac{Y-(-1)}{3} < 1\right)$$

by NSNID fact and by A(i) above

$$\approx 0.683 = 68.3\%.$$

(3) Suppose Z is $N(\mu,\sigma)$. Find the proportion of data in Z that's within 3 std devs of the mean.

Solution

We're looking for:

$$P(\mu-3\sigma < Z < \mu+3\sigma)$$

$$= P\left(\frac{\cancel{\mu}-3\sigma-\cancel{\mu}}{\sigma} < \frac{Z-\mu}{\sigma} < \frac{\cancel{\mu}+3\sigma-\cancel{\mu}}{\sigma}\right)$$

$$= P\left(-3 < \frac{Z-\mu}{\sigma} < 3\right) = 99.7\%,$$

by A(i) above.