Avd. Matematisk statistik

KTH Teknikvetenskap

Sf 2955: Computer intensive methods :
SKEW-NORMAL DISTRIBUTIONS/ Timo Koski

A random variable X is said to have a *skew-normal* distribution, if it has the density

$$f(x;\lambda) = 2\phi(x)\Phi(\lambda x), \quad -\infty < x < \infty,$$

where $-\infty < \lambda < \infty$ and $\phi(x)$ is the probability density of the standard normal distribution $N(0,1)$, or

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

and $\Phi(x)$ is the distribution function $\Phi(x) = \int_{-\infty}^{x} \phi(u)du$. We write $X \in \mathrm{SN}(\lambda)$ and note that $\mathrm{SN}(0) = N(0,1)$. We have two plots of $f(x;\lambda)$ in figure 1.

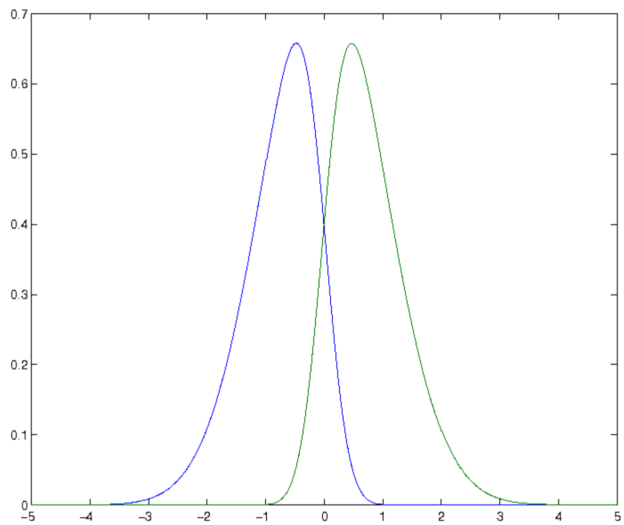
Next we write down a brief list of some of the first properties of skew-normal distributions.

1 Properties of $\mathrm{SN}(\lambda)$

1. It is to be checked that

$$\int_{-\infty}^{\infty} f(x;\lambda)dx = 1 \quad \text{for all } \lambda.$$

A hint: define $\Psi(\lambda) := \int_{-\infty}^{\infty} f(x;\lambda)dx$. Then we have $\Psi(0) = 1$ and $\frac{d}{d\lambda}\Psi(\lambda) = 0$ for all λ (check this) and thus the claim is proved.



Figur 1: The densities of SN (-3) (the left hand function graph) and SN (3) (the right hand function graph).

2. If $\lambda \to \infty$, then $f(x; \lambda)$ converges pointwise to

$$f(x) = \begin{cases} 2\phi(x) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases}$$

which is a folded normal distribution. If $\lambda \to -\infty$, then $f(x; \lambda)$ converges pointwise to

$$f(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ 2\phi(x) & \text{if } x < 0, \end{cases}$$

which is another folded normal distribution.

3.

$$X^2 \in \chi^2_1.$$

This can be seen by computing

$$P\left(X^2 \leq t\right)$$

and differentiating.

4. $$E[X] = \sqrt{\frac{2}{\pi}} \frac{\lambda}{\sqrt{1+\lambda^2}}.$$

   To see this, introduce a new auxiliary function $\Psi(\lambda) := \int_{-\infty}^{\infty} x f(x;\lambda) dx$ and find
   $$\frac{d}{d\lambda}\Psi(\lambda) = \sqrt{\frac{2}{\pi}} \frac{1}{(1+\lambda^2)^{3/2}}.$$
   Then
   $$E[X] = \int \sqrt{\frac{2}{\pi}} \frac{1}{(1+\lambda^2)^{3/2}} d\lambda + C.$$
   and constant of integration C can be determined from $\Psi(0) = 0$.

5. $$V[X] = 1 - \frac{2}{\pi} \frac{\lambda^2}{1+\lambda^2}$$

   This is easily find by noting that
   $$V[X] = E\left[X^2\right] - (E[X])^2.$$
   and the fact that $E[X^2] = 1$, as $X^2 \in \chi_1^2$.

6. **Skewness** of a random variable X is a measure of symmetry, or more precisely, the lack of symmetry of its distribution. A distribution (or data set) is symmetric if it looks the same to the left and right of the center point. Skewness κ_1 is mathematically defined as
   $$\kappa_1 = E\left[\frac{(X - E(X))^3}{\sigma^3}\right] = \frac{E(X^3) - 3E(X)\sigma^2 - (E(X))^3}{\sigma^3}$$
   The skewness of $X \in \mathrm{SN}(\lambda)$ is found to be
   $$\kappa_1 = \left(\frac{4-\pi}{2}\right) \cdot \frac{(E[X])^3}{(V[X])^{3/2}}.$$
   Hence $\lambda = 0$ implies $\kappa_1 = 0$.

7. **Kurtosis** is a measure of whether the distribution of X is peaked or flat relative to a normal distribution. High kurtosis (a.k.a. *leptokurtosis*) tends to have a distinct peak near the mean, decline rather rapidly, and

have heavy tails. Data sets with low kurtosis (a.k.a. *platykurtosis*) tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case. Kurtosis κ_2 is mathematically defined as

$$\kappa_2 = E\left[\frac{(X - E(X))^4}{\sigma^4}\right].$$

If $X \in N(m, \sigma^2)$, then the kurtosis is computed to be $= 3$. Kurtosis can in this sense be used to measure how much a distribution differs from the normal distribution. The kurtosis of $X \in \mathrm{SN}(\lambda)$ is calculated to be

$$\kappa_2 = 2(\pi - 3) \cdot \frac{(E[X])^4}{(V[X])^2}.$$

8. **Theorem 1.1** *let $X \in N(0,1)$ and $Y \in N(0,1)$ and X and Y be independent. Take a real number λ. Set*

$$Z = \begin{cases} Y, & \text{if } \lambda Y \geq X \\ -Y, & \text{if } \lambda Y < X. \end{cases}$$

Then $Z \in \mathrm{SN}(\lambda)$.

Obviously this gives a simple way of simulating random samples of $\mathrm{SN}(\lambda)$ using, e.g., Matlab^R.

Proof: The distribution function $F_Z(z)$ is

$$F_Z(z) = P(Z \leq z)$$

and by construction of Z we get that the event $Z \leq z$ is the union of two disjoint events and thus

$$P(Z \leq z) = P(\{Y \leq z\} \cap \{X < \lambda Y\}) + P(\{-Y \leq z\} \cap \{X \geq \lambda Y\}). \tag{1.1}$$

We consider the first term in the right hand side and get

$$P(\{Y \leq z\} \cap \{X < \lambda Y\}) = \int_{-\infty}^{z} P(X < \lambda y \mid Y = y) f_Y(y) dy$$

and since X och Y are independent

$$= \int_{-\infty}^{z} P(X < \lambda y) f_Y(y) dy =$$

$$= \int_{-\infty}^{z} \left(\int_{-\infty}^{\lambda y} \phi(u) du \right) f_Y(y) dy = \int_{-\infty}^{z} \Phi(\lambda y) f_Y(y) dy,$$

where $\Phi(y)$ is the distribution function of $X \in N(0,1)$. We write this as

$$= \int_{-\infty}^{z} \Phi(\lambda y) \phi(y) dy,$$

where $\phi(y)$ is the probability density of $Y \in N(0,1)$.
For the second term in the right hand side of (1.1) we have that

$$P(\{-Y \le z\} \cap \{X \ge \lambda Y\}) = P(\{Y \ge -z\} \cap \{X \ge \lambda Y\}).$$

As in the first case we get

$$P(\{Y \ge -z\} \cap \{X \ge \lambda Y\}) = \int_{-z}^{\infty} P(X \ge \lambda y \mid Y = y) f_Y(y) dy$$

$$= \int_{-z}^{\infty} P(X \ge \lambda y) \phi(y) dy$$

$$= \int_{-z}^{\infty} (1 - P(X < \lambda y)) \phi(y) dy = \int_{-z}^{\infty} (1 - \Phi(\lambda y)) \phi(y) dy.$$

The change of variable $y = -u$ yields

$$= -\int_{z}^{-\infty} (1 - \Phi(-\lambda u)) \phi(-u) du = \int_{-\infty}^{z} \Phi(\lambda u) \phi(u) du,$$

where we used an elementary rule of intergration, the symmetry $\phi(-u) = \phi(u)$ and the formula $\Phi(-\lambda u) = 1 - \Phi(\lambda u)$. By inserting this in (1.1) we get thus

$$P(Z \le z) = \int_{-\infty}^{z} \Phi(\lambda y) \phi(y) dy + \int_{-\infty}^{z} \Phi(\lambda u) \phi(u) du =$$

dvs.

$$= 2 \int_{-\infty}^{z} \Phi(\lambda y) \phi(y) dy.$$

When we differentiate this w.r.t. z we get the density

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} P(Z \le z) = 2\phi(z)\Phi(\lambda z)$$

as was claimed. ■

2 Estimation of λ

The maximum likelihood estimate of λ can be found numerically for independent samples $x_1, x_2, \ldots, x_n$.

We evoke the method of *moment estimation* using the expectation. The moment estimator is found by setting the empirical (= arithmetic) mean equal to the population mean, i.e.,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sqrt{\frac{2}{\pi}}\frac{\lambda}{\sqrt{1+\lambda^2}}$$

and then solving w.r.t. λ. This gives

$$\mid \widehat{\lambda} \mid = \sqrt{\frac{\bar{x}^2}{\frac{2}{\pi} - \bar{x}^2}}.$$

This is a plug-in estimator. For the estimator to be defined it is required that the samples are such that

$$\mid \bar{x} \mid < \sqrt{\frac{2}{\pi}},$$

and when this is not true, then we need to use some other estimator. When bootstrapping $\mid \widehat{\lambda} \mid$ we simply discard those bootstrap samples, where the condition is not satisfied.

We are in the computer demonstrations mostly taking $\lambda > 0$, and then make

$$\widehat{\lambda} = \widehat{\lambda}\,(x_1, x_2, \ldots, x_n) = \sqrt{\frac{\bar{x}^2}{\frac{2}{\pi} - \bar{x}^2}}.$$

The questions of standard error and distribution of $\widehat{\lambda}$ are hard to answer using exact analysis. Clearly, these can be addressed by bootstrapping. For this see the slide on the accompanying attachment on the course page for current information.