

Truncated normal distribution

From Wikipedia, the free encyclopedia

In probability and statistics, the **truncated normal distribution** is the probability distribution of a normally distributed random variable whose value is either bounded below or above (or both). The truncated normal distribution has wide applications in statistics and econometrics. For example, it is used to model the probabilities of the binary outcomes in the probit model and to model censored data in the Tobit model.

Contents

- 1 Definition
- 2 Moments
- 3 Simulating
- 4 See also
- 5 References

Definition

Suppose $X \sim N(\mu, \sigma^2)$ has a normal distribution and lies within the interval $X \in (a, b)$, $-\infty \leq a < b \leq \infty$. Then X conditional on $a < X < b$ has a truncated normal distribution.

Its probability density function, f , for $a \leq x \leq b$, is given by

$$f(x; \mu, \sigma, a, b) = \frac{\frac{1}{\sigma} \phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}$$

and by $f = 0$ otherwise.

Here,

$$\phi(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right)$$

is the probability density function of the standard normal distribution and $\Phi(\cdot)$ is its cumulative distribution function. There is an understanding that if $b = \infty$, then $\Phi\left(\frac{b-\mu}{\sigma}\right) = 1$, and similarly, if $a = -\infty$, then $\Phi\left(\frac{a-\mu}{\sigma}\right) = 0$.

Moments

Let $\alpha = (a - \mu)/\sigma$ and $\beta = (b - \mu)/\sigma$.

Two sided truncation:[2]

$$\mathbb{E}(X \mid a < X < b) = \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} = \mu + \sigma \frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)}$$

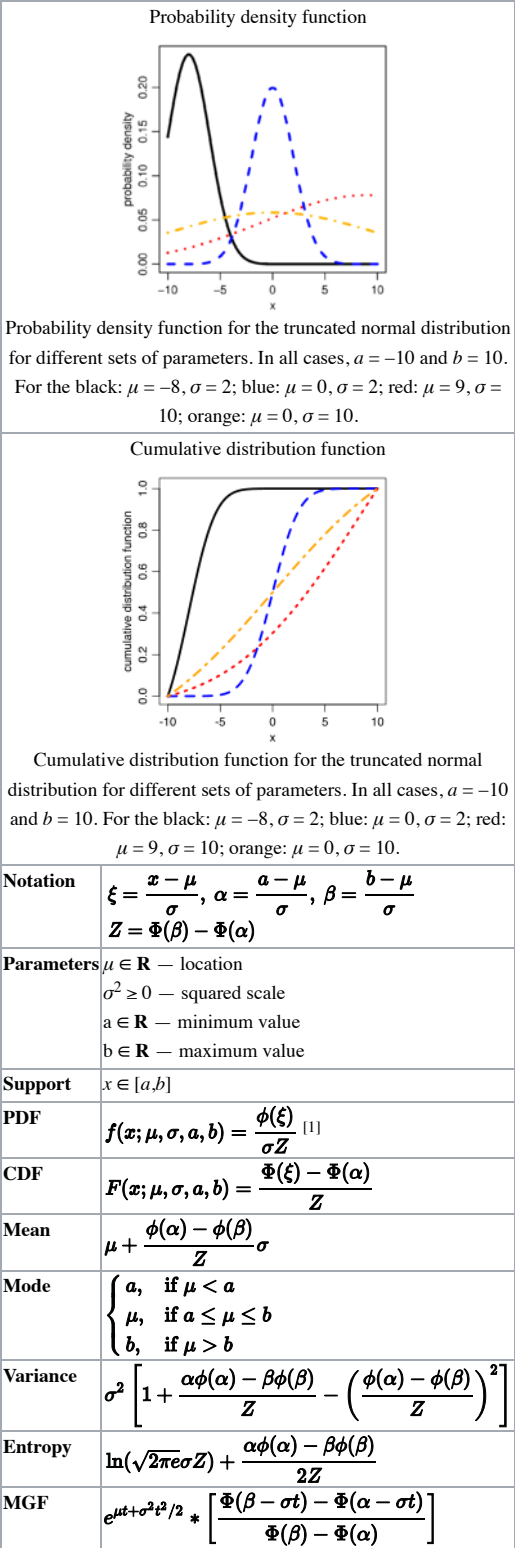
$$\text{Var}(X \mid a < X < b) = \sigma^2 \left[1 + \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} - \left(\frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \right)^2 \right] = \sigma^2 \left[1 + \frac{\alpha \phi(\alpha) - \beta \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right]$$

One sided truncation (upper tail):[3]

$$\begin{aligned} \mathbb{E}(X \mid X > a) &= \mu + \sigma \lambda(\alpha) \\ \text{Var}(X \mid X > a) &= \sigma^2 [1 - \delta(\alpha)], \end{aligned}$$

where $\alpha = (a - \mu)/\sigma$, $\lambda(\alpha) = \phi(\alpha)/[1 - \Phi(\alpha)]$ and $\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$.

One sided truncation (lower tail):



$$\mathbf{E}(X \mid X < b) = \mu - \sigma \frac{\phi(\beta)}{\Phi(\beta)}$$

$$\mathbf{Var}(X \mid X < b) = \sigma^2 \left[1 - \beta \frac{\phi(\beta)}{\Phi(\beta)} - \left(\frac{\phi(\beta)}{\Phi(\beta)} \right)^2 \right],$$

Barr and Sherrill (1999) give a simpler expression for the variance of one sided truncations. Their formula is in terms of the chi-square CDF, which is implemented in standard software libraries. Bebu and Mathew (2009) provide formulas for (generalized) confidence intervals around the truncated moments.

Differential equation

$$\left\{ \sigma^2 f'(x) + f(x)(x - \mu) = 0, f(0) = \frac{\sqrt{\frac{2}{\pi}} e^{-\frac{\mu^2}{2\sigma^2}}}{\sigma \left(\operatorname{erf}\left(\frac{\mu-a}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{\mu-b}{\sqrt{2}\sigma}\right) \right)} \right\}$$

A recursive formula

As for the non-truncated case, there is a neat recursive formula for the truncated moments. See.^[4]

Simulating

A random variate x defined as $\mathbf{x} = \Phi^{-1}(\Phi(\alpha) + U \cdot (\Phi(\beta) - \Phi(\alpha)))\sigma + \mu$ with Φ the cumulative distribution function and Φ^{-1} its inverse, U a uniform random number on $(0, 1)$, follows the distribution truncated to the range (a, b) . This is simply the inverse transform method for simulating random variables. Although one of the simplest, this method can either fail when sampling in the tail of the normal distribution,^[5] or be much too slow. Thus, in practice, one has to find alternative methods of simulation.

One such truncated normal generator (implemented in Matlab (<http://www.mathworks.com/matlabcentral/fileexchange/53180-truncated-normal-generator>) and in R (programming language) as `trandn.R` (<https://cran.r-project.org/web/packages/TruncatedNormal/>)) is based on an acceptance rejection idea due to Marsaglia.^[6] Despite the slightly suboptimal acceptance rate of Marsaglia (1964) in comparison with Robert (1995), Marsaglia's method is typically faster, because it does not require the costly numerical evaluation of the exponential function.

For more on simulating a draw from the truncated normal distribution, see Robert (1995), Lynch (2007) Section 8.1.3 (pages 200–206), Devroye (1986). The MSM (<https://cran.r-project.org/web/packages/msm/index.html>) package in R has a function, `rtnorm` (<http://rss.acs.unt.edu/Rdoc/library/msm/html/tnorm.html>), that calculates draws from a truncated normal. The `truncnorm` (<https://cran.r-project.org/web/packages/truncnorm/>) package in R also has functions to draw from a truncated normal.

Chopin (2011) proposed (arXiv (<http://arxiv.org/abs/1201.6140>)) an algorithm inspired from the Ziggurat algorithm of Marsaglia and Tsang (1984, 2000), which is usually considered as the fastest Gaussian sampler, and is also very close to Ahrens's algorithm (1995). Implementations can be found in C (http://www.crest.fr/~ckfinder/userfiles/files/Pageperso/chopin/truncnorm_20120618.tgz), C++ (<http://miv.u-strasbg.fr/mazet/rtnorm/rtnormCpp.zip>), Matlab (<http://miv.u-strasbg.fr/mazet/rtnorm/rtnormM.zip>) and Python (<http://www.christophlassner.de/blog/2013/08/12/Generation-of-Truncated-Gaussian-Samples/>).

Sampling from the *multivariate* truncated normal distribution is considerably more difficult.^[7] Exact or perfect simulation is only feasible in the case of truncation of the normal distribution to a polytope region.^[7] In more general cases, Damien and Walker (2001) introduce a general methodology for sampling truncated densities within a Gibbs sampling framework. Their algorithm introduces one latent variable and, within a Gibbs sampling framework, it is more computationally efficient than the algorithm of Robert (1995).

See also

- Normal distribution
- Truncated distribution

References

- "Lecture 4: Selection" (pdf). *web.ist.utl.pt*. Instituto Superior Técnico. November 11, 2002. p. 1. Retrieved 14 July 2015.
- Johnson, N.L., Kotz, S., Balakrishnan, N. (1994) *Continuous Univariate Distributions, Volume 1*, Wiley. ISBN 0-471-58495-9 (Section 10.1)
- Greene, William H. (2003). *Econometric Analysis (5th ed.)*. Prentice Hall. ISBN 0-13-066189-9.
- Document by Eric Orjebini, "http://www.smp.uq.edu.au/people/YoniNazarathy/teaching_projects/studentWork/EricOrjebini_TruncatedNormalMoments.pdf"
- Kroese, D. P.; Taimre, T.; Botev, Z. I. (2011). *Handbook of Monte Carlo methods*. John Wiley & Sons.
- Marsaglia, George (1964). "Generating a variable from the tail of the normal distribution". *Technometrics*. **6** (1): 101–102. doi:10.2307/1266749.
- Botev, Z. I. (2016). "The normal law under linear restrictions: simulation and estimation via minimax tilting". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. doi:10.1111/rssb.12162.
- Greene, William H. (2003). *Econometric Analysis (5th ed.)*. Prentice Hall. ISBN 0-13-066189-9.
- Norman L. Johnson and Samuel Kotz (1970). *Continuous univariate distributions-1*, chapter 13. John Wiley & Sons.
- Lynch, Scott (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer. ISBN 978-1-4419-2434-6.
- Robert, Christian P. (1995). "Simulation of truncated normal variables". *Statistics and Computing*. **5** (2): 121–125. doi:10.1007/BF00143942.
- Barr, Donald R.; Sherrill, E.Todd (1999). "Mean and variance of truncated normal distributions". *The American Statistician*. **53** (4): 357–361. doi:10.1080/00031305.1999.10474490.
- Bebu, Ionut; Mathew, Thomas (2009). "Confidence intervals for limited moments and truncated moments in normal and lognormal models". *Statistics and Probability Letters*. **79**: 375–380. doi:10.1016/j.spl.2008.09.006.
- Damien, Paul; Walker, Stephen G. (2001). "Sampling truncated normal, beta, and gamma densities". *Journal of Computational and Graphical Statistics*. **10** (2): 206–215. doi:10.1198/10618600152627906.
- Nicolas Chopin, "Fast simulation of truncated Gaussian distributions". *Statistics and Computing* **21**(2): 275-288, 2011, doi:10.1007/s11222-009-9168-1 (<http://dx.doi.org/10.1007/s11222-009-9168-1>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Truncated_normal_distribution&oldid=730570800"

Categories: Continuous distributions | Normal distribution | Probability distributions

-
- This page was last modified on 19 July 2016, at 21:32.
 - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.