# Chapter 2: Flexible probability distributions

Christophe Ley

University of Luxembourg, 2023-2024

Statistical Modelling

# Outline

# Why bother ?

If classical probability distributions have drawbacks as seen in Chapter 1, why bother at all working with parametric models ? Should we not simply adopt non-parametric approaches all the time ?

# Why bother ?

If classical probability distributions have drawbacks as seen in Chapter 1, why bother at all working with parametric models ? Should we not simply adopt non-parametric approaches all the time ?

This is a serious question. There exist many excellent non-parametric statistical methods, and machine learning methods are typically non-parametric. So again : why bother ?

# Why bother ?

If classical probability distributions have drawbacks as seen in Chapter 1, why bother at all working with parametric models ? Should we not simply adopt non-parametric approaches all the time ?

This is a serious question. There exist many excellent non-parametric statistical methods, and machine learning methods are typically non-parametric. So again : why bother ?

Probability distributions are the building blocks of statistical modelling and inference. Perhaps more basically yet importantly, they allow us to quantify the uncertainty of random phenomena and to describe the random behavior of data by means of a mathematical formula.
The latter aspect seems to be a core need in scientists, see the search for universal formulae explaining physical phenomena or epidemics such as the Covid-19 crisis where researchers strive to detect whether the growth rates are of an exponential type or approach rather a more reassuring logistic or Gompertz curve.

Describing a random phenomenon with a probability distribution is not a goal per se, but it allows inter alia calculating concrete probabilities of events and risks, making predictions and defining strategies related to the data at hand.

- A historic example is the scale of Intelligence Quotient points, which has been chosen to be a normal distribution centered at 100 and with standard deviation of 15, allowing to derive to what upper percentile a person with e.g. an IQ of 133 belongs.
- In the domain of renewable energies, the two-parameter Weibull distribution is generally adopted as model for wind speed, load and power, and thus used to determine quantities such as capacity factor and average output of a wind turbine.
- In sports, various rankings are based on well interpretable parameters of probability distributions that model match outcomes.

# Further applications

- Data analysis : often, the parameters or combinations of parameters provide information about the location, scale, skewness, kurtosis, dependence or other aspects of data shapes, and this gives the data analyst a more concrete way to describe and investigate the data at hand.
- Calculation of relevant quantities : these may be risks of exceeding a certain threshold, correlation measures such as Kendall's tau, survival functions, peak and duration of epidemics, or fan charts used by banks to quantify uncertainty, ...
- Enrichment of other statistical techniques : good probability distributions can serve as versatile basis for other statistical methods such as (quantile) regression, time series analysis, Bayesian statistics (as choices for priors), among others.

# Further applications

- Stochastic modelling : in situations where it is impossible to calculate probabilities of certain events (e.g., spread of a disease, winner of a tournament, rainfall, development of ecological systems), it is crucial to be able to simulate them repeatedly in order to approximate the true unknown probabilities. Or in order to carry out stress tests of statistical procedures !

# Famous aphorism

## All models are wrong, but some are useful

Generally attributed to the famous statistician George Box.

# Famous aphorism

## All models are wrong, but some are useful

Generally attributed to the famous statistician George Box.

We shall in the sequel discuss flexible probability distributions and what we require from them in order to be “useful”.

# What properties should a good flexible model possess ?

# What properties should a good flexible model possess ?

- versatility : as the word “flexible” suggests, such a model should ideally be able to exhibit as many distinct shapes as possible and, consequently, be more robust to misspecifications than simple models.
- tractability : the density should be of a tractable form and amenable to calculations.
- interpretability : the number of parameters should be as small as possible and the parameters should bear clear interpretations in order to infer conclusions about the underlying population from which the data were taken.
- straightforward parameter estimation : a correct parameter estimation procedure is the basis for the subsequent calculation of quantities such as the risk of exceeding a certain threshold value and for statistical inference.

# What properties should a good flexible model possess ?

- data generating mechanism : this desideratum has two goals : 1) we should be able to readily simulate new data from the model in order to produce large-scale stochastic simulations and predictions (e.g., about the spread of a disease), and 2) when related to a nice stochastic representation, we may be able to naturally link a particular model to data for which we can trace back their real generation process.
- testability and model reduction : natural goodness-of-fit tests can be defined for the flexible model, and it should nest well-known sub-models, as this permits model reduction.

# Outline

# Transforming the normal

There exist many distinct options in the literature to build flexible distributions coping with asymmetry and tailweight. A classical choice is to start from the normal distribution, which has a location and a scale parameter, and to change it in such a way that either skewness, either tailweight, or both are added, typically via a skewness and/or tailweight parameter.

There exist models where one parameter influences both skewness and tailweight, and this is fine. From an interpretation point of view it is, however, easier with one parameter for each role.

We shall discuss here in detail one model that adds a skewness parameter : the famous skew-normal distribution.

# The skew-normal

Originally proposed by De Helguero (1908) but really popularized by Azzalini (1985), the skew-normal distribution has a density given by

$$f_{SN}(x;\mu,\sigma,\lambda) = \frac{2}{\sigma}\phi(\sigma^{-1}(x-\mu))\Phi\left(\lambda\sigma^{-1}(x-\mu)\right)$$

where $\mu \in \mathbb{R}$ is the location, $\sigma \in \mathbb{R}_0^+$ is the scale and $\lambda \in \mathbb{R}$ is the skewness parameter. Clearly $\lambda = 0$ yields the normal distribution as special case.

A random variable $X$ following this density is written $X \sim SN(\mu,\sigma,\lambda)$ and $SN(\lambda)$ for the standard skew-normal.

Let us first ensure that this construction leads to a well-defined density. We shall to this end prove the following result, which has turned out to play a major role in what is now known as symmetry modulation.

**Lemma**

*Let $f$ be a density symmetric about 0, and $G$ a non-negative function such that $G(-x) = 1 - G(x)$. Then $2f(x)G(\lambda x)$ is a density function for any real $\lambda$.*

Let us prove this result together.

# Some immediate properties

- When $\lambda \to \infty$, we retrieve the half-normal density.
- When $Z \sim SN(\lambda)$, then $-Z \sim SN(-\lambda)$.
- The skew-normal density is strictly unimodal, in other words, $\log f_{SN}(x;\lambda)$ is a concave function.
- When $Z \sim SN(\lambda)$, then $Z^2 \sim \chi^2_1$.

# Moments

Let $Z \sim SN(\lambda)$. Then its moment generating function corresponds to $M_Z(t) = 2\exp(t^2/2)\Phi(\delta t)$ with $\delta = \lambda/\sqrt{1+\lambda^2}$. Some algebra then leads to the following important quantities :

$$E(Z) = b\delta, \quad Var(Z) = 1 - (b\delta)^2, \quad \gamma_2(Z) = 2(\pi - 3)\left(\frac{(E(Z))^2}{Var(Z)}\right)^2$$

and

$$\gamma_1(Z) = \frac{1}{2}(4-\pi)\text{sign}(\lambda)\left(\frac{(E(Z))^2}{Var(Z)}\right)^{3/2},$$

where $b = \sqrt{2/\pi}$, and $\gamma_1$ and $\gamma_2$ respectively denote the third and fourth standardized cumulants.

# Shape of the skew-normal density

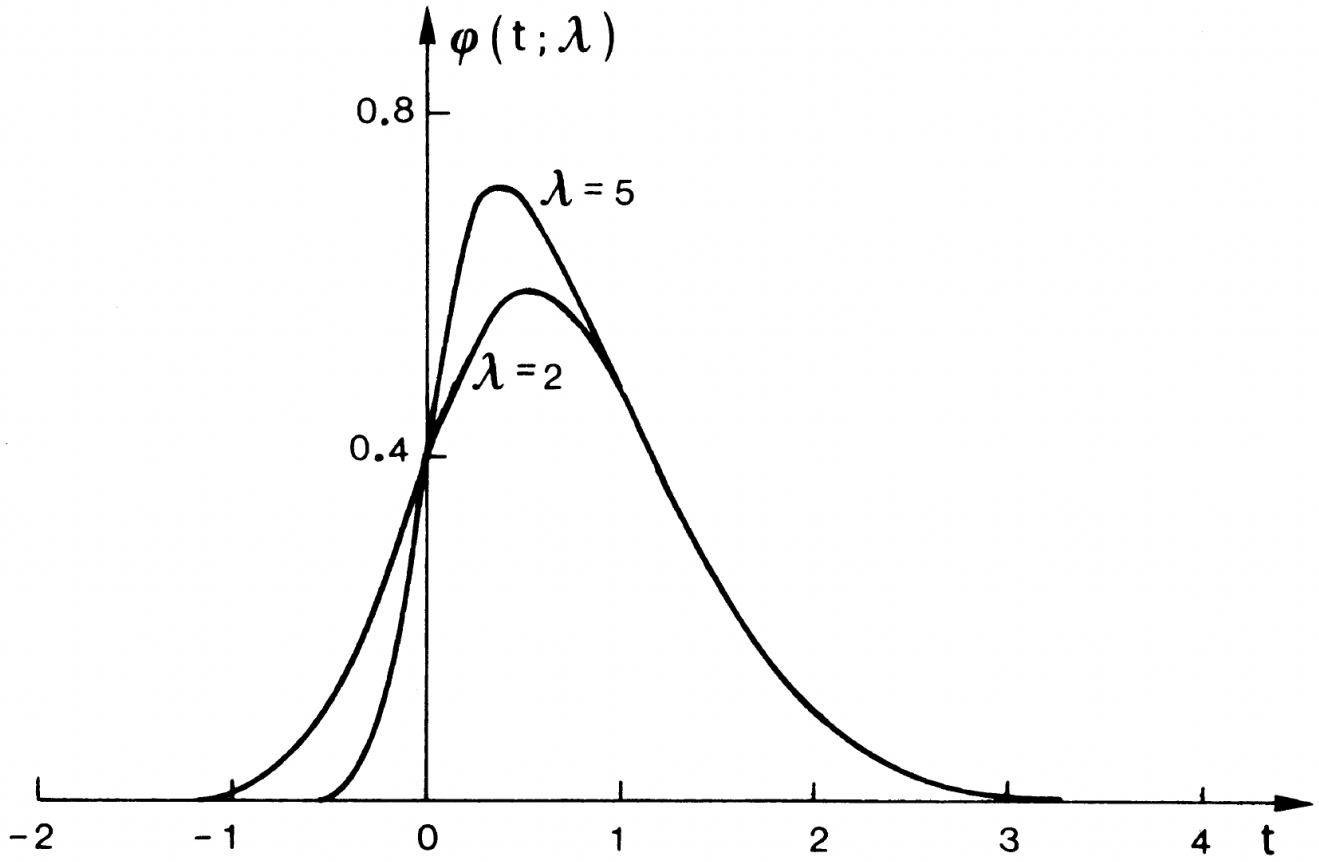



*Fig. 1.* The density functions SN(2) and SN(5).

Taken from Azzalini (1985).

# Random number generation

There exist several stochastic representations for the skew-normal. This is one of its main attractive features.

Mechanism 1 : Let $\lambda > 0$, and $X$ and $Y$ be independent and identically distributed (iid) random variables from a $\mathcal{N}(0, 1)$ population. Then the random variable $Z = Y \mid \lambda Y > X$ follows $SN(\lambda)$. This example is to be proved.

Mechanism 2. Let $(U_0, U_1)$ be a bivariate normal random vector with standardized marginals and correlation $\rho$, then $\max(U_0, U_1) \sim SN(\lambda)$ where $\lambda = \sqrt{(1-\rho)/(1+\rho)}$. This example is not to be proved.

Mechanism 3 : Let $(U_0, U_1)$ be a bivariate normal random vector. Then the distribution of $U_1 | U_0 > 0$ is skew-normal and this bears a nice interpretation concerning university exam grades. This example is not to be proved.

# A concrete example taken from Ley (2015, Journal de la SFdS

**Example 1.** *Stochastic Frontier Analysis (SFA) is concerned with the specification and estimation of a frontier production function, e.g., for firms. Economic modelling for SFA has been initiated simultaneously by Aigner et al. (1977) and Meeusen and van den Broeck (1977) and can be formulated as follows:*

$$Y = f(\mathbf{x};\boldsymbol{\beta}) + \varepsilon, \tag{1.1}$$

*where $Y$ is the observed scalar output, the production frontier $f$ depends on the input $\mathbf{x} \in \mathbb{R}^k$ and some parameter $\boldsymbol{\beta} \in \mathbb{R}^p$ to be estimated, and $\varepsilon$ is the error term. This term itself can be expressed as*

$$\varepsilon = V - U \tag{1.2}$$

*where $V$ is a random shock, assumed to be symmetric, and $U$, independent of $V$, is the random non-negative technical (in-)efficiency component inherent to each firm. Now, the structure (1.2) clearly shows that the composed error term $\varepsilon$ cannot follow a normal distribution, since it is the sum of a symmetric term $(V)$ and a negative term $(-U)$, leading to skewness in the error term.*

# Statistical Inference

In practice, one will often work with the family of distributions generated by the linear transformation

$$Y=\lambda_1+\lambda_2 Z \qquad (\lambda_2>0). \tag{6}$$

The Fisher information for the parameter $(\lambda_1, \lambda_2, \lambda)$ is easily computed, obtaining

$$I_\lambda = \begin{pmatrix} (1+\lambda^2 a_0)/\lambda_2^2 & \left(E(Z)\,\frac{1+2\lambda^2}{1+\lambda^2}+\lambda^2 a_1\right)\Big/\lambda_2^2 & \left(\frac{b}{(1+\lambda^2)^{3/2}}-\lambda a_1\right)\Big/\lambda_2 \\ \left(E(Z)\frac{1+2\lambda^2}{1+\lambda^2}+\lambda^2 a_1\right)\Big/\lambda_2^2 & (2+\lambda^2 a_2)/\lambda_2^2 & -\lambda a_2/\lambda_2 \\ \left(\frac{b}{(1+\lambda^2)^{3/2}}-\lambda a_1\right)\Big/\lambda_2 & -\lambda a_2/\lambda_2 & a_2 \end{pmatrix}$$

where

$$a_k=a_k(\lambda)=E\left\{Z^k\left(\frac{\phi(\lambda Z)}{\Phi(\lambda Z)}\right)^2\right\} \qquad (k=0,\,1,\,2),$$

which has to be evaluated numerically.

Taken from Azzalini (1985)

In the skew-normal case, solutions have been proposed, such as reparameterizations. The research on this Fisher information singularity has contributed to the fame of the skew-normal.

In the skew-normal case, solutions have been proposed, such as reparameterizations. The research on this Fisher information singularity has contributed to the fame of the skew-normal.

A multivariate extension is readily doable, and we shall see it later in this chapter.

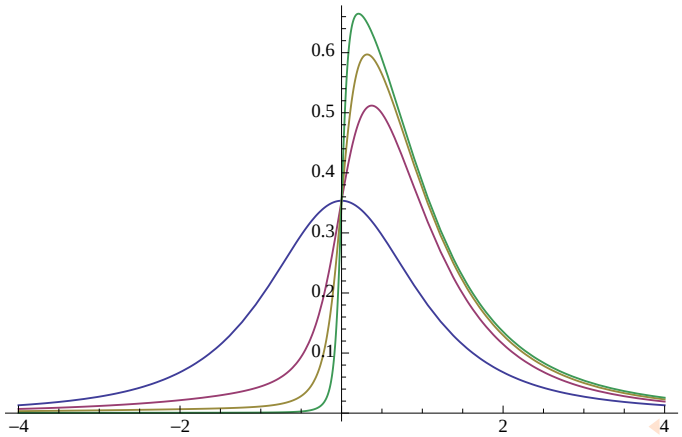
Let us go back to the list of desirable properties and discuss how "good" the skew-normal is as flexible model.

# The skew-$t$

Extensions to other distributions are also quite straightforward. A particularly interesting case is the skew-$t$ distribution with density

$$2t(x;\nu)T\left(\lambda x\sqrt{\frac{1+\nu}{x^2+\nu}};\nu+1\right)$$

where $t(\cdot;m)$ and $T(\cdot;m)$ stand for the density and distribution function of the Student $t$ distribution with $m>0$ degrees of freedom (note that the Cauchy distribution corresponds to $t(\cdot;1)$).
Thanks to the presence of the tailweight parameter, the skew-$t$ allows modelling both skewness and tailweight.



# Outline

# The Akaike Information Criterion

The goal of the Akaike Information Criterion (AIC) is to estimate the information loss when assuming a certain model is responsible for the data generating process. In other words, if we assume that the true process follows from a density $f_{\text{true}}$, then the AIC measures the fit of some density $f$ by estimating its distance to the unknown $f_{\text{true}}$.

AIC thus estimates the relative amount of information lost by a given model : the less information a model loses, the higher the quality of that model.

The AIC deals with the trade-off between the goodness-of-fit of a model and its simplicity, thus dealing both with over-fitting and under-fitting.

# Origins

The AIC is due to the Japanese statistician Hirotugu Akaike who first presented his idea during a symposium in 1971. The first formal publication appeared in 1974. This paper is today cited nearly 58000 times on googlescholar (early March 2022).

He founded his criterion on information theory. He measured the amount of information loss between $f$ and $f_{\text{true}}$ via the Kullback-Leibler divergence :

$$D_{\text{KL}}(f_{\text{true}}||f) = \int f_{\text{true}}(x) \log(f_{\text{true}}(x)/f(x))dx.$$

Of course we cannot calculate this quantity since we do not know $f_{\text{true}}$. Akaike showed by mathematical arguments that one can estimate with his AIC the model-dependent estimation loss for each model. This estimate is asymptotically valid ; hence for small sample sizes, the AIC might be less good.

# The formula

For a density $f = p_{\theta_m}$ with $k_m$ parameters that need to be estimated, the AIC is given by

$$\mathrm{AIC}(p_{\theta_m}) = 2k_m - 2\ell_{n,m}(\hat{\theta}_m).$$

# The formula

For a density $f = p_{\theta_m}$ with $k_m$ parameters that need to be estimated, the AIC is given by

$$\mathrm{AIC}(p_{\theta_m}) = 2k_m - 2\ell_{n,m}(\hat{\theta}_m).$$

There exists a small-sample correction, the AICc defined as

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2k_m^2 + 2k_m}{n - k_m - 1}.$$

Note that the AIC reveals nothing about the absolute quality of a model, only the quality relative to other models. Thus, if all the candidate models fit poorly, the AIC will choose the least bad one but it might be far from giving a good fit. This is why one should accompany the AIC by a goodness-of-fit (GOF) test.

# An alternative : the BIC

There exists a famous alternative to the AIC, namely the Bayesian Information Criterion (BIC). We will not delve into Bayesian details here and simply give its formula :

$$\mathrm{BIC}(p_{\theta_m}) = \log(n)k_m - 2\ell_{n,m}(\hat{\theta}_m).$$

We thus see that the penalty term has changed, giving more weight to the number of parameters when the sample size increases, compared to the AIC.

The same criticism as for AIC appeals to the BIC, underlining the need for GOF tests.

# Normality tests

Before discussing general GOF tests, we shall now see two procedures that allow to consider the problem $H_0$ : the data follow a normal distribution against $H_A$ : they are not normally distributed.

# The Jarque-Bera test

The Jarque-Bera test uses two characteristics of the normal curve : symmetry and tail-weight. These quantities are measured through the coefficients

$$b_3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{3/2}}$$

and

$$b_4 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{2}}.$$

In case of the normal distribution : $b_3 = 0$ and $b_4 = 3$. This is the idea underpinning the Jarque-Bera test statistic

$$JB = \frac{n}{6}b_3^2 + \frac{n}{24}(b_4 - 3)^2$$

which, under the null hypothesis, follows a $\chi^2_2$ distribution. When do we hence reject the null hypothesis ?

# The Lilliefors test

This test relies on a comparison between the normal cumulative distribution function and the empirical cumulative distribution function of the data :

$$LF = \sup_x \left| \Phi(x; \hat{\mu}, \hat{\sigma}) - \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(X_i \leq x) \right|$$

where $\hat{\mu}$ and $\hat{\sigma}$ are estimates of the parameters $\mu$ and $\sigma$. Here $\mathcal{I}$ is an indicator function.

The normality hypothesis is rejected for large values of $LF$ ; the critical values can be found in tables, but mostly are built-in in the software programs.

# Chi-square goodness-of-fit test

We have seen tests for normality. Fine. But how to test for other distributional assumptions ?

We shall now see a well-known answer to this question : the chi-square goodness-of-fit test.

As an illustration, we will investigate whether the number of hail storms in Manitoba, Canada, follows a Poisson distribution with $\lambda = 1.2$.

http ://www.producer.com/2015/10/summer-of-hailstorms-across-manitoba/

## Hailstorms in Manitoba

| $X = \#$ hail storms per year | $\#$ years where $X$ is observed | Poisson probabilities | $\#$ years where $X$ is expected |
|---|---|---|---|
| 0 | 10 | 0.301 | 10.5 |
| 1 | 13 | 0.361 | 12.6 |
| 2 | 8 | 0.217 | 7.6 |
| 3 | 3 | 0.087 | 3.1 |
| 4 | 1 | 0.026 | 0.9 |
| 5+ | 0 | 0.008 | 0.3 |
| Total | 35 | 1.000 | 35.0 |

Table – Number of expected hail storms per year in Manitoba under a Poisson distribution.

Modus operandi :

$H_0$ : The data follow a Poisson(1.2) distribution.
$H_A$ : They don't !

Chi-square test statistic :

$$\sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} \sim \chi^2_{k-1} \quad |H_0$$

where $N_j$ is the number of observations in class $j = 1, \ldots, k$, $k$ the number of distinct values taken by the data, $n$ the total sample size, and $p_1, \ldots, p_k$ the theoretical Poisson(1.2) probabilities.

We reject $H_0$ at asymptotic level $\alpha$ when $X \notin [0; \chi^2_{k-1,1-\alpha}]$.

For the hail storms in Manitoba, we get the following test statistic :

$$\begin{aligned} & \frac{(10-10.5)^2}{10.5} + \frac{(13-12.6)^2}{12.6} + \frac{(8-7.6)^2}{7.6} + \\ & \frac{(3-3.1)^2}{3.1} + \frac{(1-0.9)^2}{0.9} + \frac{(0-0.3)^2}{0.3} \\ = \quad & 0.3719 \end{aligned}$$

The critical value is $\chi^2_{6-1;\ 0.95} = 11.07$ : we do not have enough evidence to reject $H_0$ !

# The Kolmogorov-Smirnov test

The KS test is based on the test statistic

$$D_n = \sup_x |F_n(x) - F(x)|$$

where $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(x_i \leq x)$ and $F$ is the candidate distribution whose fit we wish to test.

The Glivenko-Cantelli theorem states that, if the sample comes from $F$, then $D_n$ converges to 0 almost surely. The KS test therefore considers $\sqrt{n}D_n$ whose asymptotic distribution is the complicated Kolmogorov distribution, linked to a Brownian bridge. This distribution is typically built-in in the classical softwares.

# The Cramér-von Mises test

The Cramér-von Mises test is an alternative to the KS test and based on

$$CM_n = \int_{\mathbb{R}} \left(F_n(x) - F(x)\right)^2 dF(x).$$

The test statistic $nCM_n$ has a complicated distribution that is, again, contained in most classical softwares.

# The Cramér-von Mises test

The Cramér-von Mises test is an alternative to the KS test and based on

$$CM_n = \int_{\mathbb{R}} \left(F_n(x) - F(x)\right)^2 dF(x).$$

The test statistic $nCM_n$ has a complicated distribution that is, again, contained in most classical softwares.

There exist obviously more GOF tests, such as Anderson-Darling, Kuiper, or tests based on kernel density estimation.

# COM-Poisson Distribution Properties

- Simulation studies demonstrate COM-Poisson flexibility
  - Table II assesses goodness of fit on simulated data of size 500

**Table II**. True model parameters versus model estimates (and associated goodness-of-fit p-values provided in parentheses) for various assumed distributions

| | Estimated parameter | | | |
|---|---|---|---|---|
| True distribution | Poisson | Geometric | COM-Poisson | |
| Poisson($\lambda=10$) | $\lambda=9.986$ (0.9436) | $p=0.091$ (0.0000) | $\lambda=10.244$ (0.8904) | $\nu=1.011$ |
| Geometric($p=0.2$) | $\lambda=3.862$ (0.0000) | $p=0.206$ (0.6220) | $\lambda=0.794$ (0.6220) | $\nu=0.000$ |
| COM-Poisson($\lambda=10,\ \nu=5$) | $\lambda=1.184$ (0.0000) | $p=0.458$ (0.0000) | $\lambda=12.223$ (0.4284) | $\nu=5.267$ |
| COM-Poisson($\lambda=3,\ \nu=0.5$) | $\lambda=9.510$ (0.0000) | $p=0.095$ (0.0000) | $\lambda=3.157$ (0.6604) | $\nu=0.522$ |

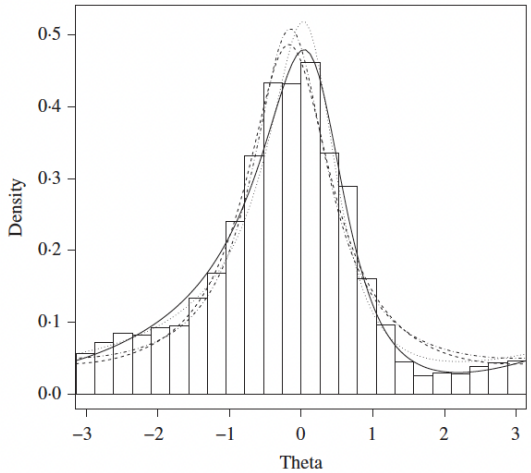



Fig. 3. Histogram of the gun crime data with maximum likelihood fits of the full model (4) (solid), the three-parameter symmetric submodel (6) (dashed), the three-parameter asymmetric submodel (7) (dotted), and the wrapped Cauchy submodel (dot-dashed).

Table 1. *Maximum likelihood estimates of the parameters, with standard errors in square brackets, the maximized loglikelihood, $\ell_{\max}$, and the values of Akaike Information Criterion, AIC, and Bayesian Information Criterion, BIC, for the full model (4) and four of its submodels fitted to the gun crime data*

| Model | $\hat{\mu}$ | $\hat{\gamma}$ | $\hat{\alpha}_2$ | $\hat{\beta}_2$ | $\ell_{\max}$ | AIC | BIC |
|---|---|---|---|---|---|---|---|
| Full model (4) | −0·384 [0·018] | 0·547 [0·009] | 0·233 [0·011] | 0·115 [0·008] | −44741·00 | 89490·00 | 89520·68 |
| Three-parameter symmetric (6) | −0·302 [0·018] | 0·543 [0·009] | 0·257 [0·010] | (0) | −44992·29 | 89990·58 | 90013·59 |
| Three-parameter asymmetric (7) | −0·341 [0·017] | 0·519 [0·008] | (0·269) | 0·095 [0·008] | −44841·44 | 89688·88 | 89711·89 |
| Wrapped Cauchy | −0·286 [0·017] | 0·524 [0·008] | (0·274) | (0) | −45025·35 | 90054·70 | 90070·04 |
| Cardioid | −0·510 [0·025] | 0·432 [0·006] | (0) | (0) | −46059·42 | 92122·84 | 92138·18 |

# Kato and Jones (2015)

| | Model | | | |
|---|---|---|---|---|
| Parameter | Normal | Normal tails | Symmetric | Family (2) |
| $\xi$ | 39.24 | 24.66 | 40.94 | -65.26 |
| $\eta_\delta$ | 20.20 | 17.88 | 24.89 | 0.85 |
| $\delta$ | 1 | 1 | 19263.4 | 4.24 |
| $\epsilon$ | 0 | 0.52 | 0 | 17.09 |
| Diagnostics | | | | |
| $l_{max}$ | -504.39 | -502.50 | -497.72 | -494.93 |
| A | 1012.78 | 1011.00 | 1001.44 | 997.86 |
| B | 1018.25 | 1019.21 | 1009.65 | 1008.80 |
| $p$-value | 0.031 | 0.006 | 0.311 | 0.253 |

Table 2: *Parameter estimates for the fits to the ice floe snow depth data of, reading from right to left, family (2) and its symmetric,* $\epsilon = 0$*, normal-tailed,* $\delta = 1$ *and normal,* $\delta = 1, \epsilon = 0$*, sub-models. The maximised log–likelihood,* $l_{max}$*, Akaike Information Criterion, A, Bayesian Information criterion, B, and p-value for the chi–squared goodness-of-fit test are included as fit diagnostics.*

Jones and Pewsey (2009)

# Outline

# What is different in higher dimensions ?

In $k > 1$ dimensions, the data rapidly become much more complex than in dimension 1.

Indeed, skewness and tailweight can be different along the various dimensions. Moreover, we now have to take the dependence structure into account. This can quickly result in mathematically difficult, even intractable, probability distributions, and this is not what we want in general...

# Forms of symmetry

Speaking of skewness : there already exist many distinct types of symmetry in higher dimensions :

- Spherical symmetry : $\mathbf{X}$ is spherically symmetric about the origin iff $\mathbf{X} \stackrel{d}{=} \mathbf{O}\mathbf{X}$ for any $k \times k$ rotation matrix $\mathbf{O}$.
- Elliptical symmetry : $\mathbf{X}$ is elliptically symmetric with $k \times k$ scatter matrix $\mathbf{\Sigma}$ about the origin iff $\mathbf{\Sigma}^{-1/2}\mathbf{X}$ is spherically symmetric
- Central symmetry : $\mathbf{X}$ is centrally symmetric about the origin iff $\mathbf{X} \stackrel{d}{=} -\mathbf{X}$

And there exist further forms of symmetry ...

# A general overview

We shall now discover a structured overview of flexible probability distributions on $\mathbb{R}^k$ via the written document (paper)

**Comparison and classification of flexible distributions for multivariate skew and heavy-tailed data**