Chapter 1
Background

Seminar *Spatio-temporal dependence*, 07.02.2011

Benedikt Gräler
Institute for Geoinformatics
University of Muenster

Outline

1 Probability Theory

- Distributions
- Random Variables
- Extreme Value Distributions
- cdf-Transform
- n-variate Distributions
- Practical

ifgi
Institute for Geoinformatics
University of Münster

Outline

1 Probability Theory

- Distributions
- Random Variables
- Extreme Value Distributions
- cdf-Transform
- n-variate Distributions
- Practical

2 Statistics

- Estimation
- GOF
- Practical

Outline

Probability Theory

In *Probability Theory* we design *experiments* which underly some randomness and ask for the probability of certain outcomes.

Example (throwing a six-sided die)

The set of all possible values is $\Omega := \{1, \ldots, 6\}$. We assume, there is no number to occur more likely than any other. Thus, we design a *probability measure* $\mathbb{P}$, which assigns a probability of $\frac{1}{6}$ to each possible event in Ω. Thus, its *density* is the constant function $f(\omega) \equiv \frac{1}{6}$ on Ω.

Distributions

These probability measures are, in a more general sense, *distributions*. There is an unlimited number of possible distributions and families of distributions.

Example

- Normal distribution $N(\mu, \sigma^2)$ on $\mathbb{R}$ with mean $\mu \in \mathbb{R}$, standard deviation $\sigma \in (0, \infty)$ and density $f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(x-\mu)^2}{2\sigma^2})$
- Uniform distribution $U(a, b)$ on $(a, b) \subset \mathbb{R}$ (or $[a, b]$) with density $f_U(x) = \frac{1}{b-a}$
- Exponential distribution $Exp(\lambda)$ on $[0, \infty)$ with parameter $\lambda \in (0, \infty)$ and density $f_E(x) = \lambda \exp(-\lambda x)$

pdf & cdf

Each distribution has besides its density f (*pdf* $=$ probability distribution function) a cumulative distribution function F (*cdf*) that is for continuous distributions

$$F(x) := \int_{-\infty}^{x} f(t)dt$$

and in case of discrete probability distributions

$$F(x) := \sum_{\omega \in \Omega, \omega \leq x} f(\omega).$$

Properties of the cdf

We derive the following properties from the design of a cdf F:

- F is monotone non-decreasing
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to +\infty} F(x) = 1$
- F is right-continuous: $\lim_{x \to +x_0} F(x) = F(x_0)$

A pdf or cdf uniquely defines a probability distribution.

distributions in R

distribution	R command
$\beta(a, b)$	beta
$C(a, b)$	cauchy
χ^2	chisq
Exponential	exp
F(m; n)	f
Γ	gamma
$U(a, b)$	unif
logistic	logis
lognormal	lnorm
$N(\mu, \sigma^2)$	norm
t	t
$B(n, p)$	binom
geometric	geom
hypergeom.	hyper
$NB(\alpha, \theta)$	nbinom
Poisson	pois

Random Variables

Formally speaking, a (real valued) *random variable* Z is a (measurable) function $Z : \Omega \to \mathbb{R}$.

Practically speaking, a *random variable* Z describes some property which is subject to some randomness.

We usually deal with a *spatial random field* $\mathcal{Z}$ over some spatial region S. A spatial random field is a set $\mathcal{Z} := \{Z(s) | s \in S\}$ of (spatial) random variables $Z(s)$ at different locations $s \in S$.

These $Z(s)$ take for some random event $(\omega \in \Omega)$ a value in their domain (e.g. range of temperatures). Their behaviour *may* change for different locations $s \in S$ (*non-stationary* vs. *stationary* spatial random fields).

Extreme Value Distributions

A special class of probability distributions are the *extreme value distributions* (*evd*). A evd is characterized by its cdf G through the fact that

$$\lim_{n \to \infty} G^n = G.$$

Where G^n is the cdf of $\max(X_1, \ldots, X_n)$ and the X_i are independent copies of random variables following the distribution induced by G.

One can show that there are only three families of extreme value distributions: *Gumbel*, *Fréchet* and *Weibull*. These can even be parametrized as single family (*generalized extreme value distribution*).

ifgi
Institute for Geoinformatics
University of Münster

Properties of evd

The remarkable difference is in the tails of the distributions.

Extreme Value Distributions posses *heavy tails*. That is, even small regions very far from the mean of the distribution (multiple standard deviations) have a considerable positive probability.

ifgi
Institute for Geoinformatics
University of Münster

Normal vs. Gumbel

Comparing the probabilities of the standard Normal and Gumbel distribution both with mean 0 and standard deviation 1 the ratios for the first 10 multiples of the standard deviation are:

x	Gumbel/Normal
1	0.6
2	1
3	3
4	29
5	746
6	51e+3
7	9e+6
8	4e+9
9	7e+12
10	25e+15
⋮	⋮

Gumbel

The Gumbel evd is defined on $\mathbb{R}$ with parameters $\mu \in \mathbb{R}$ and $\beta \in (0, \infty)$. Its density is given by

$$f_G(x) = \frac{e^{-(x-\mu)/\beta}}{\beta} \cdot \exp(-e^{-(x-\mu)/\beta})$$

and its cumulative distribution function is

$$F_G(x) = \exp(-e^{-(x-\mu)/\beta}).$$

Fréchet

The Fréchet evd is defined on $(0, \infty)$ with parameters $\alpha \in (0, \infty)$, $m \in \mathbb{R}$ and $s \in (0, \infty)$. Its density is given by

$$f_F(x) = \alpha \left(\frac{x - m}{s} \right)^{-(1+\alpha)} \exp\left(- \left(\frac{x - m}{s} \right)^{-\alpha} \right)$$

and its cumulative distribution function is

$$F_F(x) = \exp\left(- \left(\frac{x - m}{s} \right)^{-\alpha} \right).$$

Weibull

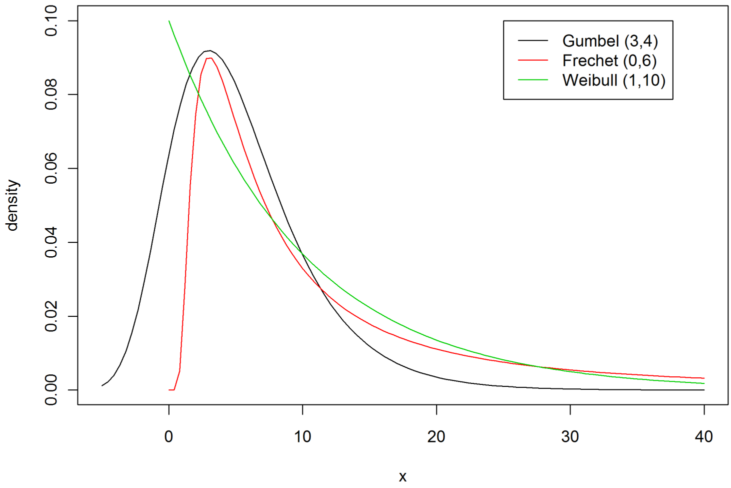
The Weibull evd is defined on $[0, \infty)$ with parameters $\lambda \in (0, \infty)$ and $k \in (0, \infty)$. Its density is given by

$$f_W(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} \cdot \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$$

and its cumulative distribution function is

$$F_W(x) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^k\right).$$

The three Extreme Value Distributions



cdf-Transform

We will need the following theorem later on:

Theorem

For a random variable X following any distribution D with cdf F_D the transformed variable $F_D(X)$ follows the standard uniform distribution $U(0,1)$.

ifgi
Institute for Geoinformatics
University of Münster

from univariate to n-variate Distributions

There are several ways of extending univariate distributions to n-variate distributions (explicit, conditional, copulas, . . .). Consider for example the n-variate normal distribution:

Example

We define new parameters: the mean vector $\boldsymbol{\mu} \in \mathbb{R}^k$ and the covariance matrix $\Sigma \in \mathbb{R}^k \times \mathbb{R}^k$ (symmetric and positive-definite). The multivariate density is given by

$$f_N(\boldsymbol{x}) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right).$$

The Joint cdf

The *joint cdf* of some n-variate distribution $\mathbb{P}_n$ is given by:

$$H(x_1, \ldots, x_n) = \mathbb{P}_n(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

In the case of bivariate distributions it is the probability of the lower left rectangle of the point (x_1, x_2) (i.e. $(-\infty, x_1) \times (-\infty, x_2)$).

Tasks

- play around with some distributions (i.e. evd) by comparing densities for different parameters of one family and between families
- find parameters for the Gumbel distribution such that its mean and standard deviation are (approximately) the same as for the standard normal distribution.
- empirically validate the theorem on the cdf-Transform

Statistics

In (geo)statistics we usually observe a *sample*/*realization* $\mathbf{Z} = \big(z(s_1), \ldots, z(s_n)\big)$ of some spatial random field $\mathcal{Z} = \{Z(s_i) | 1 \leq i \leq n\}$ and want to derive the underlying probability distribution.

Knowing the underlying probabilistic model enables us to explain the observed processes, to run simulations and to perform forecasts.

Single Characteristic Values I

Usually we start with (descriptive) estimates of single values like *mean*, *variance* or *skewness* of a given sample.

In some cases, these (moment) estimates completely determine the *parameters* of a distribution (but not the distribution itself!). This estimation procedure is called *method of moments*.

Single Characteristic Values II

Example

The parameter λ of the exponential distribution $Exp(\lambda)$ can be estimated by the mean $\hat{\mu}$ of the sample $\mathbf{Z}$ through $\hat{\lambda}(\mathbf{Z}) = \hat{\mu}^{-1}$.

The skewness of a sample may give advise in selecting a suitable distribution before one estimates the parameters.

Method of Moments

The empirical (according to the sample $\mathbf{Z} = (z_1, \ldots, z_n)$) and theoretical (of the distribution Z) moments of different order are compared. This leads to a set of equations which can then be solved for the distributions parametrs:

$$\mathbb{E}(Z) = \frac{1}{n}\sum_{i=1}^{n} z_i$$

$$\mathbb{E}(Z^2) = \frac{1}{n}\sum_{i=1}^{n} z_i^2$$

$$\mathrm{Var}(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 - \Big(\frac{1}{n}\sum_{i=1}^{n} z_i\Big)^2$$

$$\mathbb{E}(Z^3) = \frac{1}{n}\sum_{i=1}^{n} z_i^3$$

Maximum Likelihood

The *maximum likelihood* approach grounds on the assumption that the drawn sample $\mathbf{Z}$ is a very likely one. Thus, the parameter(s) $\theta \in \Theta$ are chosen for which $\mathbf{Z}$ becomes the *most likely* sample to occur:

$$\hat{\theta}_M(\mathbf{Z}) = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} f_\theta(z)$$

or equivalently

$$\hat{\theta}_L(\mathbf{Z}) = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log\big(f_\theta(z)\big)$$

Bayes Estimator

We add an *a priori* belief $p : \Theta \to [0, 1]$ to the parameter space Θ (no clue might mean uniform distribution over all parameters). The *Bayes estimator* minimizes the mean square error (MSE, or any other appropriate risk). Thus θ becomes a random variable.

For the MSE the Bayes estimator $\hat{\theta}_B(\mathbf{Z})$ is the mean of the *a posteriori* distribution of θ and given by

$$\hat{\theta}_B(\mathbf{Z}) = \mathbb{E}(\theta|\mathbf{Z}) = \int_{\Theta} \vartheta f(\vartheta|\mathbf{Z}) d\vartheta$$

where

$$f(\vartheta|\mathbf{Z}) =_{Bayes} \frac{p(\vartheta) f_\vartheta(\mathbf{Z})}{\int_{\Theta} p(t) f_t(\mathbf{Z}) dt}.$$

Tools in R

- `mean`, `var`, `sd`, `summary`
- histograms: `hist(data, freq=F)`
- overlaying functions: `curve(dnorm(x), add=T)`
- empirical cdf: `ecdf(data)` returns a step function
- quantile-quantile plots: `qqplot` compares quantiles of two samples
- calculating quantiles: `quantile(data, prob)`
- log-likelihood estimation: `mle(minuslogl)` minimizes the `minuslogl` function
- estimating a non parametric density: `density(data, bw, kernel, ...)`

Fitting Extreme Value Distributions

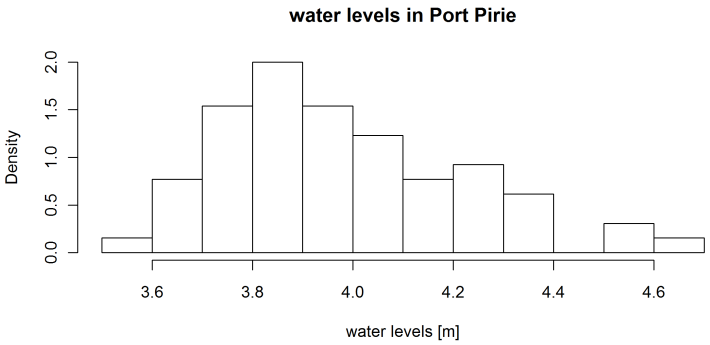
ifgi
Institute for Geoinformatics
University of Münster

We can either assume that the sample is drawn from an evd alone or a convex combination of distributions. An evd can be fitted in the same way as *usual* distributions, but there are some more tools.

- the POT-method (points over threshold) helps to select a threshold and to identify the maxima (gpd)
- annual block method: select the daily/weekly/monthly/yearly maxima as sample (gev)

Fitting a evd Using POT I

We use the dataset `portpirie` (package: `evd`) containing the annual maximum sea levels from 1923 to 1987.



Fitting a evd Using POT II

The *mean residual life function*

$$mrl(u) := \mathbb{E}(Z - u | Z > u) = \frac{\mathbb{E}\big((Z_i - u)\mathbf{1}_{Z_i > i}\big)}{\mathbb{P}(Z_i > u)}$$

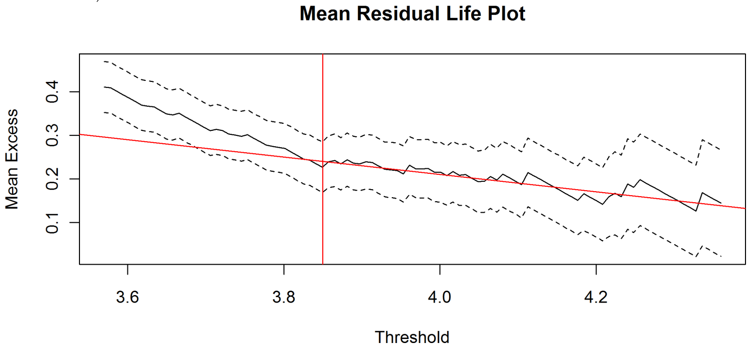
is linear in u for *generalized pareto distributions* (a joint parametrization of evd for exceedances).

Plotting $mrl(u)$ for all possible values lets us choose an appropriate threshold. The slope b of the plot can be used to calculate the *shape* parameter γ by $\hat{\gamma} = b/(1+b)$.

The parameters *scale* σ and *location* μ can be derived from a qq-plot of theoretical quantiles of a gpd with $\gamma = \hat{\gamma}$, $\sigma = 1$ and $\mu = 0$ against the sample.

Fitting a evd Using POT III

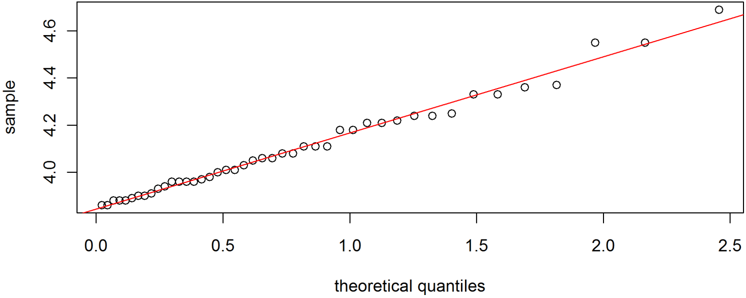
We chose a threshold of 3.85 m and fitted a line with slope $b = -0,2$.




```
mrlplot(portpirie)
abline(v=3.85,col="red")
abline(1.01,-0.2,col="red")
```

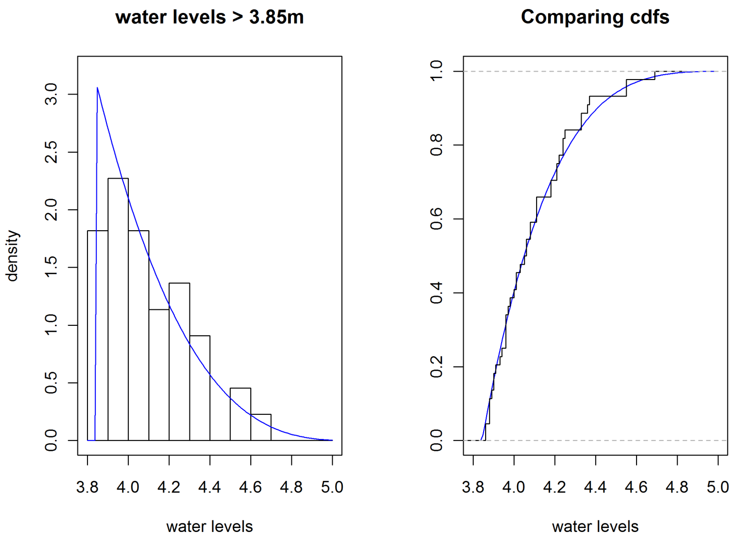
Fitting a evd Using POT IV

A linear model gave $\hat{\sigma} \approx 0.32$ and $\hat{\mu} \approx 3.84$.




```
Call:
lm(formula = sort(portpirie[portpirie > 3.85]) ~
               qgpd((1:44)/45, shape = -0.25))
Coefficients: (Intercept)   slope
                  3.8431  0.3231
```

Fitting a evd Using POT V



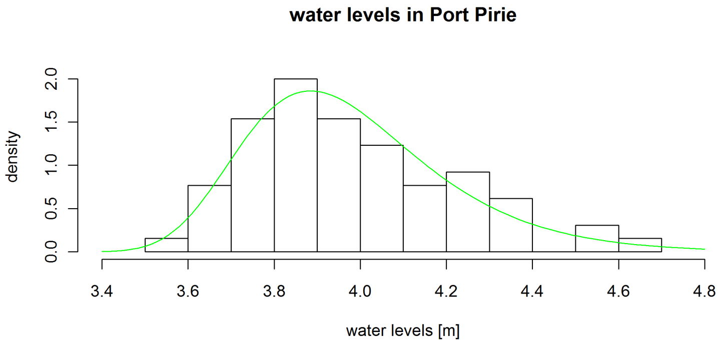
Fitting a evd Using Anual Blocks I

The data set portpirie actually is aquired as anual maxima water levels. Thus, we might try to fit a *generalized extreme value distribuitopn* (*gev*) right away.

```
> fgev(portpirie)
Call: fgev(x = portpirie)
Deviance: -8.678117

Estimates
     loc     scale     shape
 3.87475   0.19805  -0.05012

Standard Errors
    loc     scale    shape
0.02793  0.02025  0.09826

Optimization Information
  Convergence: successful
  Function Evaluations: 31
  Gradient Evaluations: 8
```

Fitting a evd Using Anual Blocks II

A visual comparison shows a quite reasonable fit:



Goodness of Fit Testing

Most techniques will give an estimate no matter how far fetched the underlying concept might be. *Goodness of fit tests* (*GOF*) are used to evaluate the fit.

Depending on the complexity of the model, a simple test-statistic might do, or a simulation needs to be run in order to compare the behavior of the fitted probability distribution to the one sample.

ifgi
Institute for Geoinformatics
University of Münster

Comparing Distributions I

The *Kolmogorov-Smirnov test* (*KS-test*) compares the ecdf with the fitted cdf by calculating its biggest difference. An alternative is the *Cramér-von-Mises test* (*CvM-test*) which considers the integral over the squared difference of two cdfs.

Example

In case of the fitted gpd we get:

```
> ks.test(exce,"pgpd",loc=3.8431,scale=0.3231,shape=-0.2)

        One-sample Kolmogorov-Smirnov test

data:  exce
D = 0.0753, p-value = 0.9644
alternative hypothesis: two-sided
```

Thus, our hypothesis (the sample was drawn from gpd with the estimated parameters) can not be rejected for any common significance level.

Comparing Distributions II

Example

In case of the automatic fit of the gev we get:

```
> ks.test(portpirie, "pgev", loc=3.87475, scale=0.19805,
                                              shape=-0.05012)

        One-sample Kolmogorov-Smirnov test

data:  portpirie
D = 0.0606, p-value = 0.9706
alternative hypothesis: two-sided
```

Thus, our hypothesis (the sample was drawn from gev with the estimated parameters) can not be rejected for any common significance level.

GOF by simulation

In cases where the underlying distribution used to calculate critical values/p-values is unknown a bootstrapping procedure can be used:

1. estimate θ from the sample $\mathbf{Z}$ through one of the given estimators
2. compute the test-statistic s_0 (e.g. CvM-test-statistic)
3. simulate a sample $\bar{\mathbf{Z}}$ from the assumed distribution of the same size as the original one
4. estimate $\bar{\theta}$ from $\bar{\mathbf{Z}}$ through the same estimator as above
5. repeat the steps 3 and 4 for a large integer $N > 100$ and compute each test-statistic $s_i, 1 \leq i \leq N$
6. the p-value can be approximated by $\#\{1 \leq i \leq N \mid s_i > s_o\}/N$.

Tasks

- fit some distributions (i.e. evd) to given samples (meuse, or use `data()` to look for more) using the methods of moments and the maximum likelihood estimator
- fit a distribution to the meuse dataset/your dataset
- implement a Bayes estimator in `R`
- conduct a goodness of fit test by simulation

- Löwe, M., Mathematische Statistik http://wwwmath.uni-muenster.de/statistik/loewe/mathstatistik.pdfl
- Löwe, M., Extremwert Theorie, http://wwwmath.uni-muenster.de/statistik/loewe/extrem.pdf