Lecture 1

BACKGROUND ON EXTREME VALUE THEORY WITH EMPHASIS ON CLIMATE APPLICATIONS

Richard W. Katz

Institute for Mathematics Applied to Geosciences
National Center for Atmospheric Research
Boulder, CO USA

email: `rwk@ucar.edu`

Home page: `www.isse.ucar.edu/staff/katz`

Lecture: `...staff/katz/docs/pdf/ubalect1.pdf`

Quote

"***It seems that the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.***"

Emil Gumbel

Outline

(1) Traditional Statistical Methods

(2) Rationale for Extreme Value Analysis

(3) Extremal Types Theorem

(4) Block Maxima Approach

(5) Tails of Distributions

(6) Choice of Threshold

(7) De-Clustering

(8) Point Process / Peaks Over Threshold

(9) Risk Communication (Under Stationarity)

(1) Traditional Statistical Methods

- **Fit models/distributions to all data**

-- Even if primary focus is on extremes

- **Statistical theory for averages**

-- Ubiquitous role of normal distribution

-- Central Limit Theorem

- **Robustness**

-- Avoid sensitivity to extremes
(outliers / contamination)

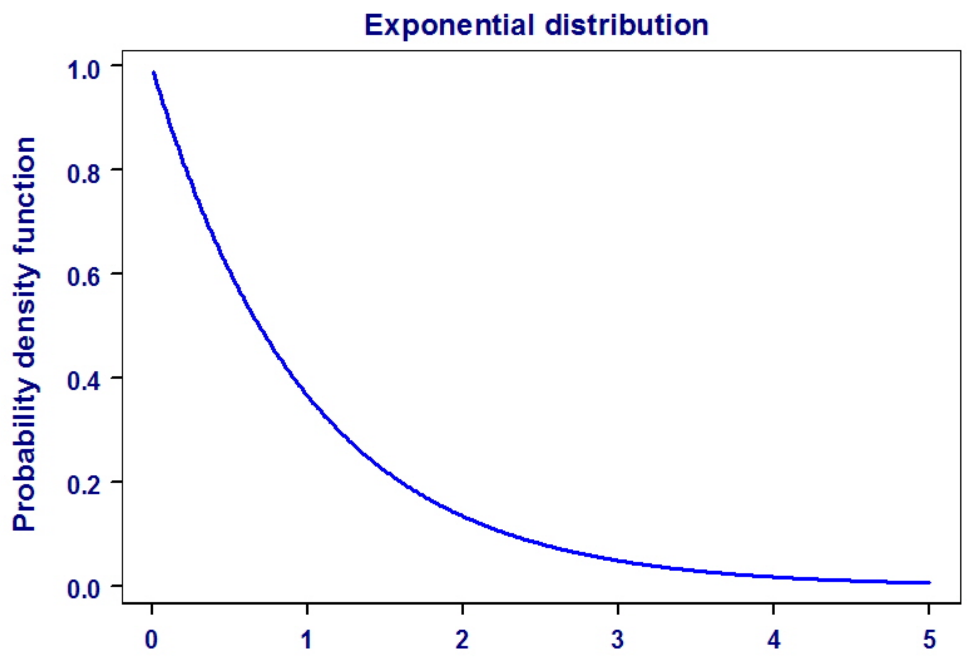
- **Nonparametric Alternatives**

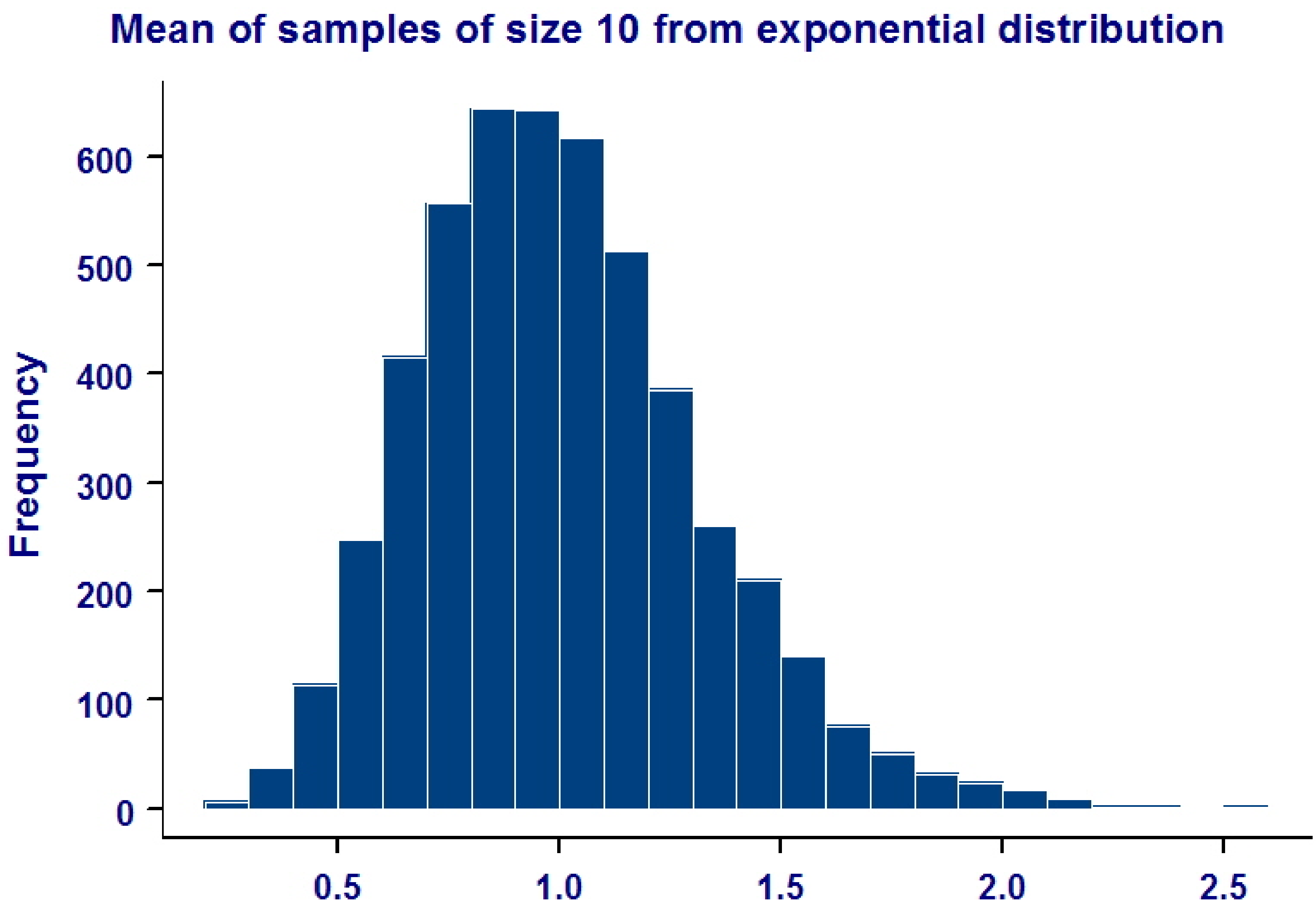
-- Kernel density estimation
Ok for center of data (but for extremes?)

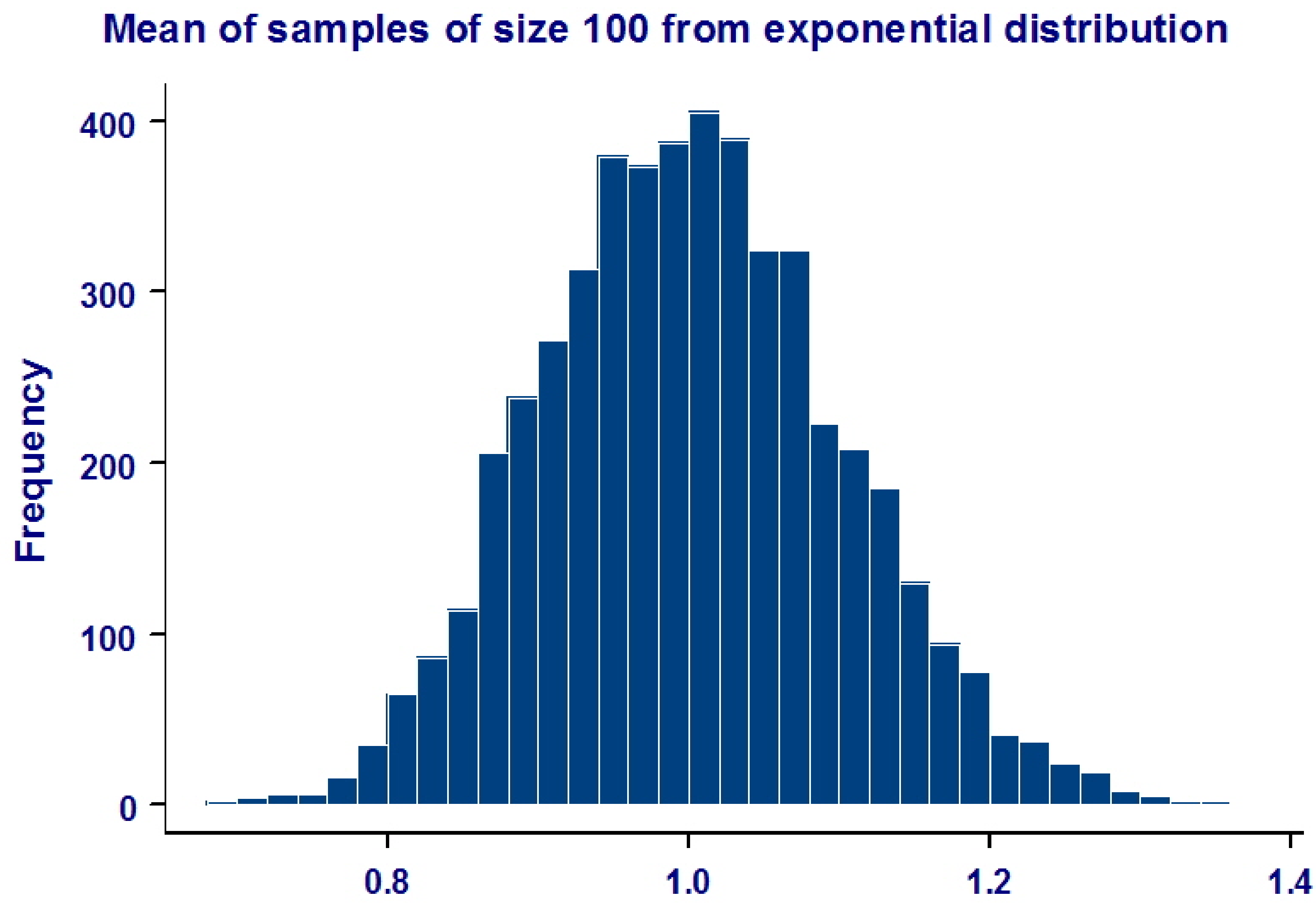
-- Resampling
Fails for maxima
Fails for heavy-tailed distributions

- **Conduct sampling experiment**

-- Draw random samples from exponential distribution (rate = 1) and calculate mean for each sample




Mean of samples of size 10 from exponential distribution
Frequency
600
500
400
300
200
100
0
0.5
1.0
1.5
2.0
2.5


Mean of samples of size 100 from exponential distribution
Frequency
0
100
200
300
400
0.8
1.0
1.2
1.4

(2) Rationale for Extreme Value Analysis

- **Limited information about extremes**

-- Exploit what theory is available

- **More robust/flexible approach**

-- Tail behavior of standard distributions is too restrictive

Statistical theory indicates possibility of "heavy" tails

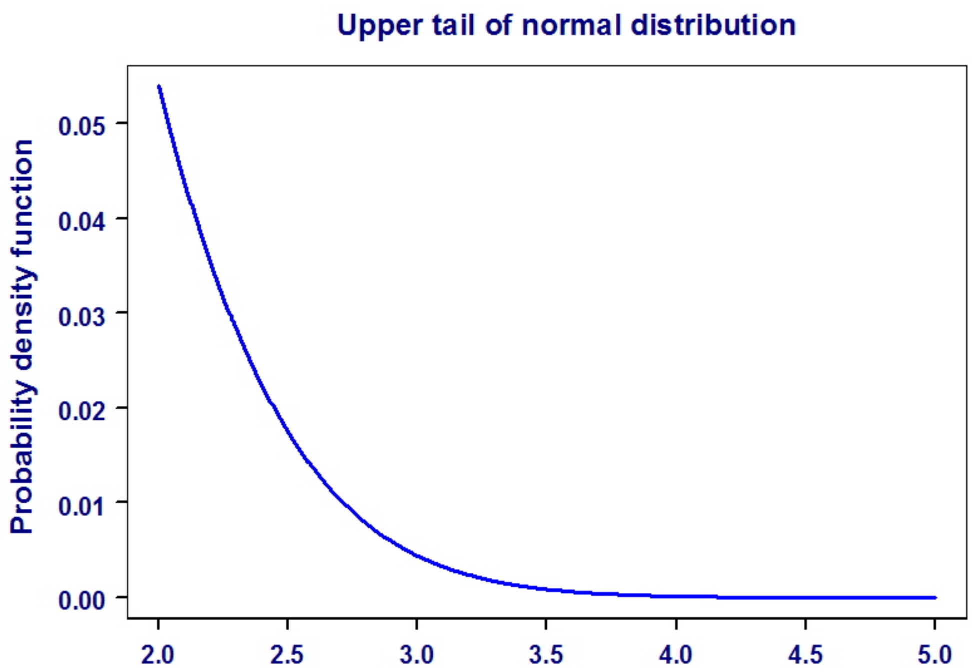
Data suggest evidence of "heavy" tails

Conventional distributions have "light" tails

-- Example

Let X have standard normal distribution with density function φ

For large x, $\Pr\{X > x\} \approx \varphi(x) / x$



- **Statistical behavior of extremes**

-- Effectively no role for normal distribution

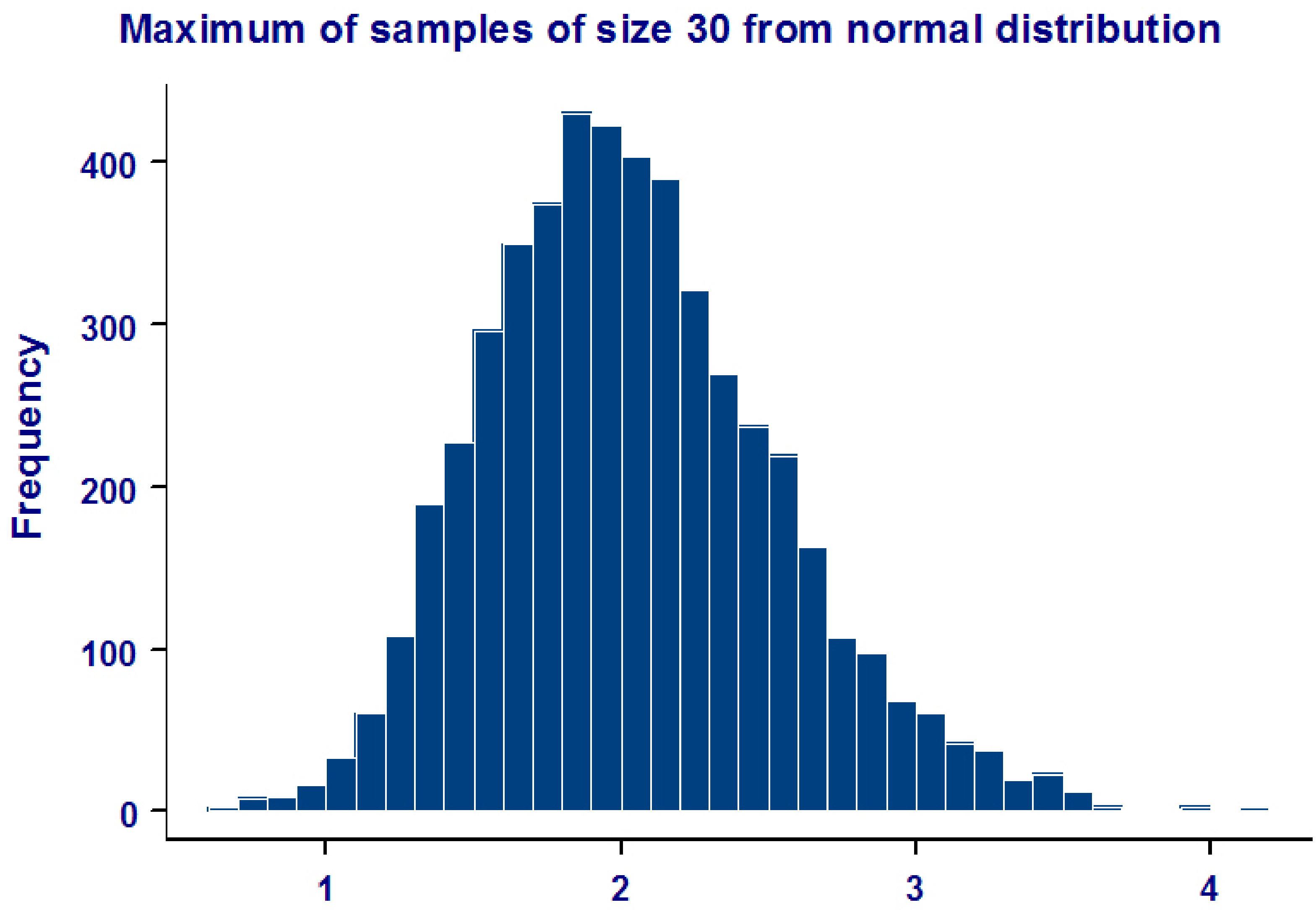
-- What form of distribution(s) instead?

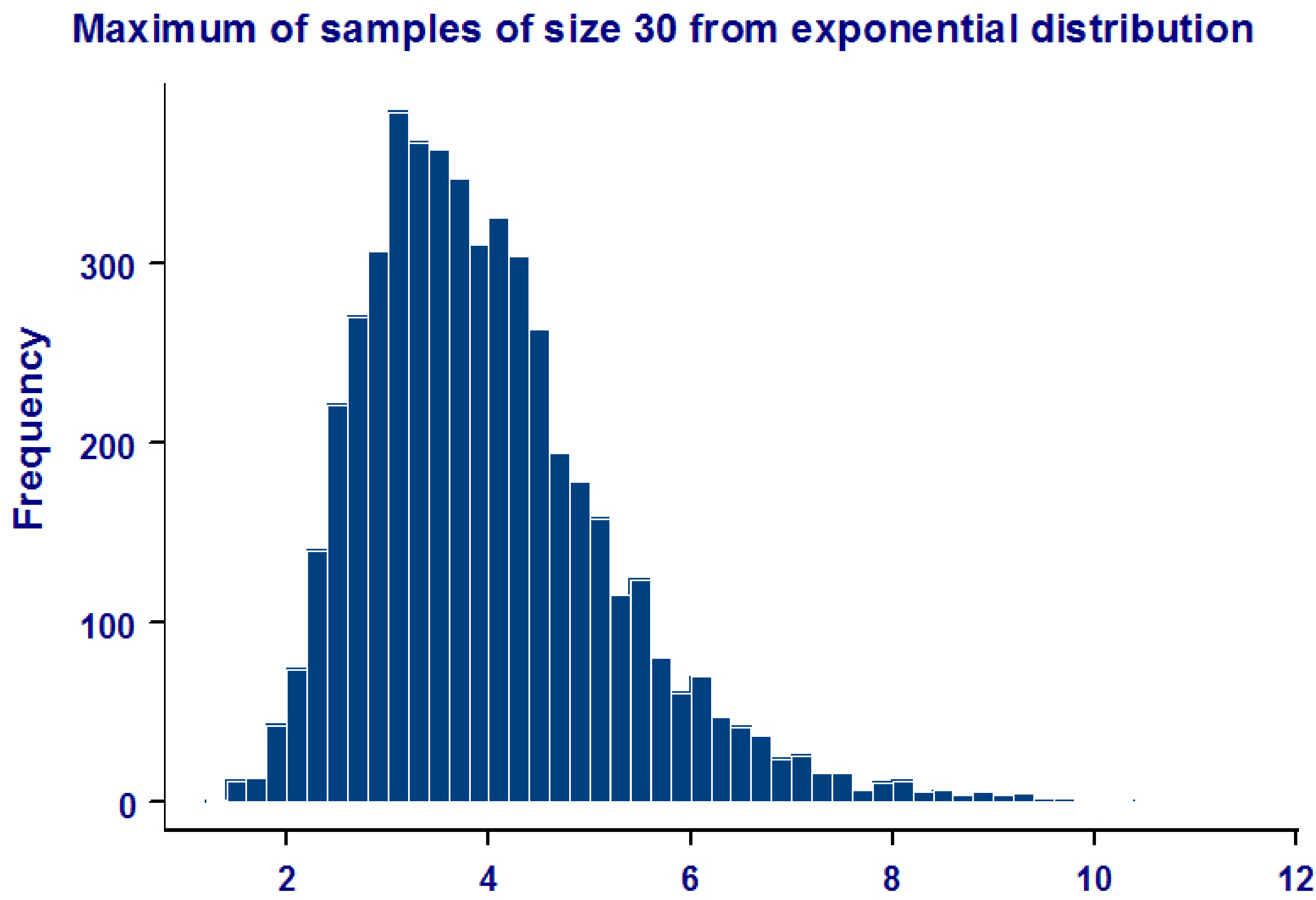
- **Conduct another sampling experiment**

-- Calculate largest value of random sample (instead of mean)

(i) Normal distribution (mean = 0, st. dev. = 1)

(ii) Exponential distribution (rate = 1)


Maximum of samples of size 30 from normal distribution
Frequency
0
100
200
300
400
1
2
3
4


Maximum of samples of size 30 from exponential distribution
Frequency
0
100
200
300
2
4
6
8
10
12

(3) Extremal Types Theorem

- **"Sum stability"**

-- Central Limit Theorem

$X_1, X_2, \ldots, X_T$ independent & identically distributed

Sum $S_T = X_1 + X_2 + \ldots + X_T$

approximately normally distributed for large T
(even when dist. of X_t's far from normal)

-- Normal distribution for X_t's

Then sum (or mean) is exactly normally distributed

- "Max stability"

-- Want to find distribution(s) for which maximum has same form as original sample

-- Extremal Types Theorem

$X_1, X_2, \ldots, X_T$ (independent & identically distributed)

Distribution of $M_T = \max\{X_1, X_2, \ldots, X_T\}$ is approximately *generalized extreme value* (GEV) for large T:

$$\Pr\{M_T \leq x\} \approx \exp\left\{-[1 + \xi\,(x - \mu)/\sigma]^{-1/\xi}\right\},\ 1 + \xi\,(x - \mu)/\sigma > 0$$

μ location parameter (depends on T)

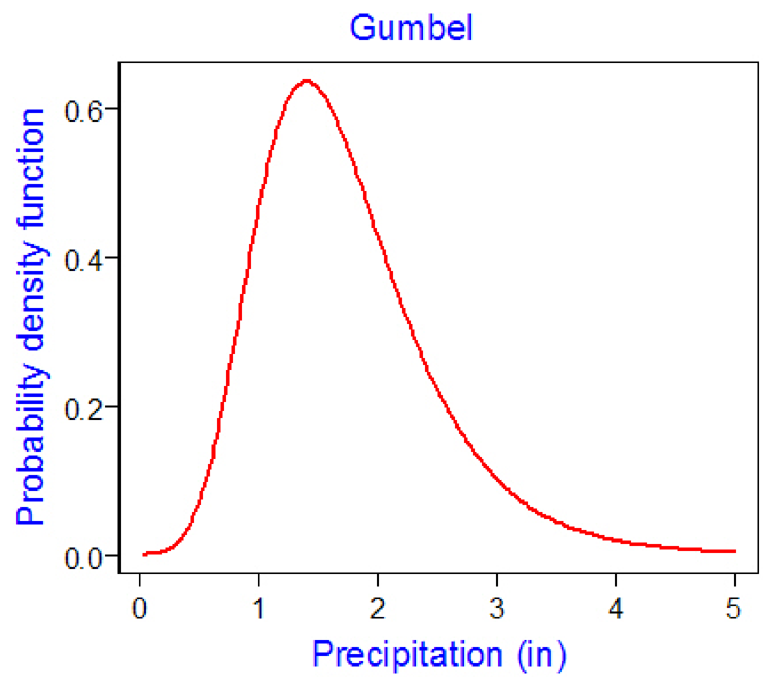
$\sigma > 0$ scale parameter (depends on T)

ξ shape parameter

(i) $\xi = 0$ (*Gumbel type*, limit as $\xi \to 0$)

“Light” tail

“Domain of attraction” for many common distributions (e. g., normal, lognormal, exponential, gamma)

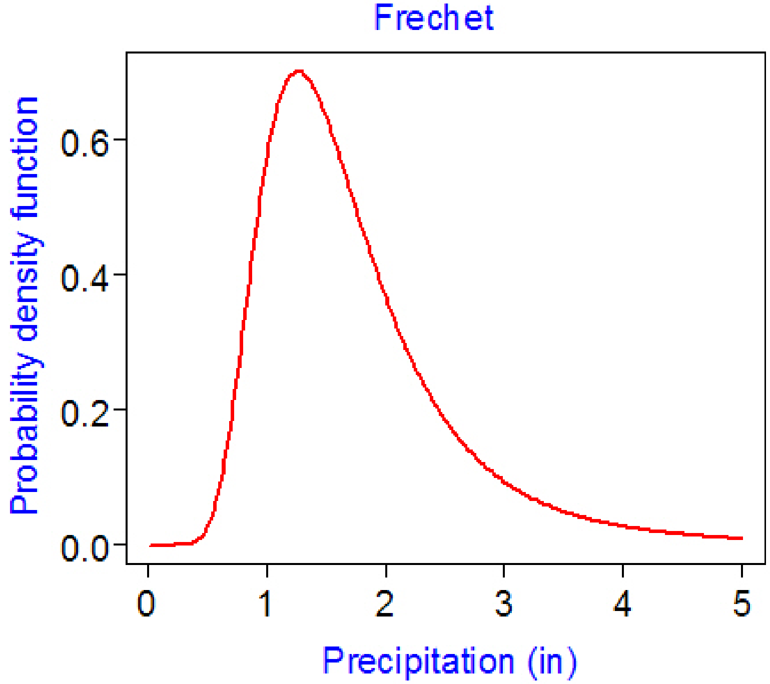


(ii) $\xi > 0$ (*Fréchet type*)

"Heavy" tail with $E(X^r) = \infty$ if $r \geq 1/\xi$

(e. g., infinite variance if $\xi \geq 1/2$)

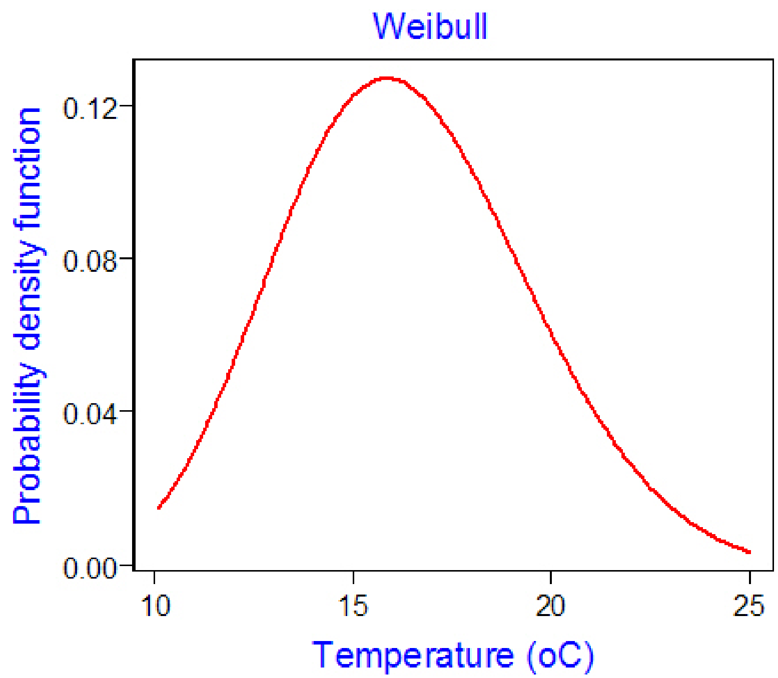
Fits precipitation, streamflow, economic impacts



(iii) $\xi < 0$ (*Weibull type*)

Bounded tail [$x < \mu + \sigma / (-\xi)$]

Fits temperature, wind speed, sea level



(4) Block Maxima Approach

- **GEV distribution**

-- Fit directly to maxima

(e. g., annual maximum of daily prec. amount or highest temperature over given year)

- **Parameter estimation**

-- Maximum likelihood estimation (MLE)

Iterative numerical procedure

- **Fort Collins daily precipitation amount**

-- Fort Collins, CO, USA

Time series of daily precipitation amount (in), 1900-1999

Semi-arid region

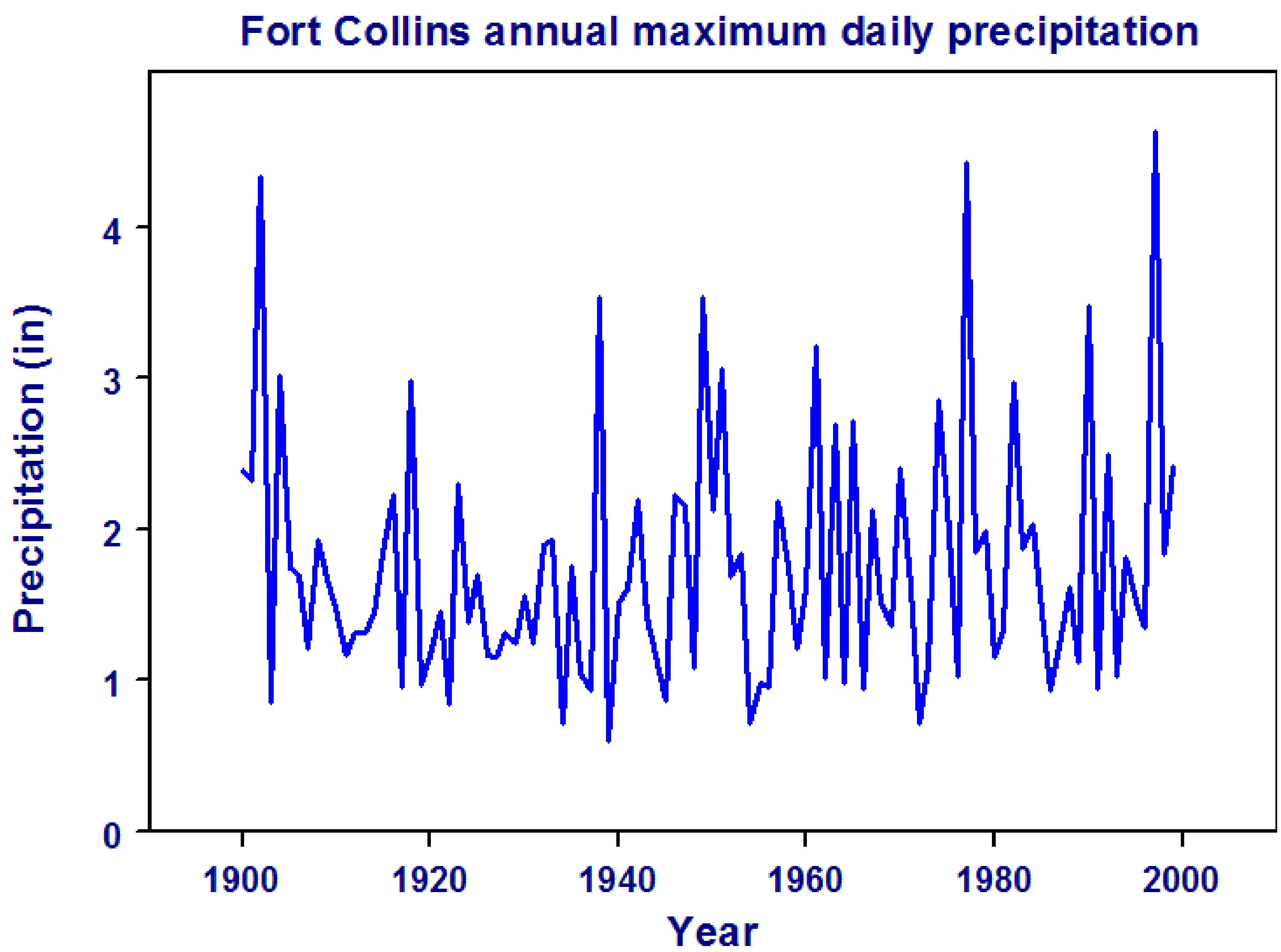
Marked annual cycle in precipitation

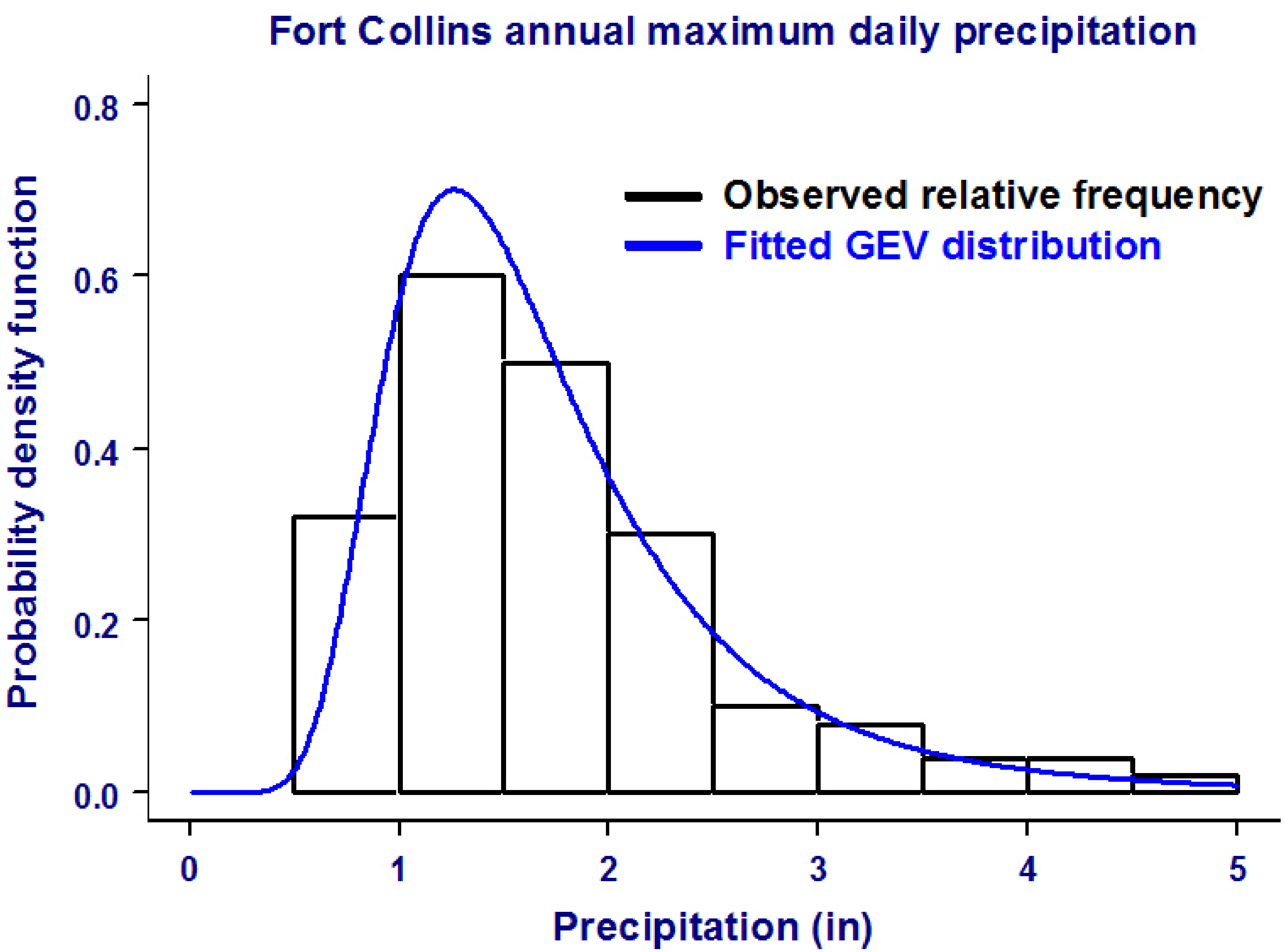
(wettest in late spring/early summer, driest in winter)

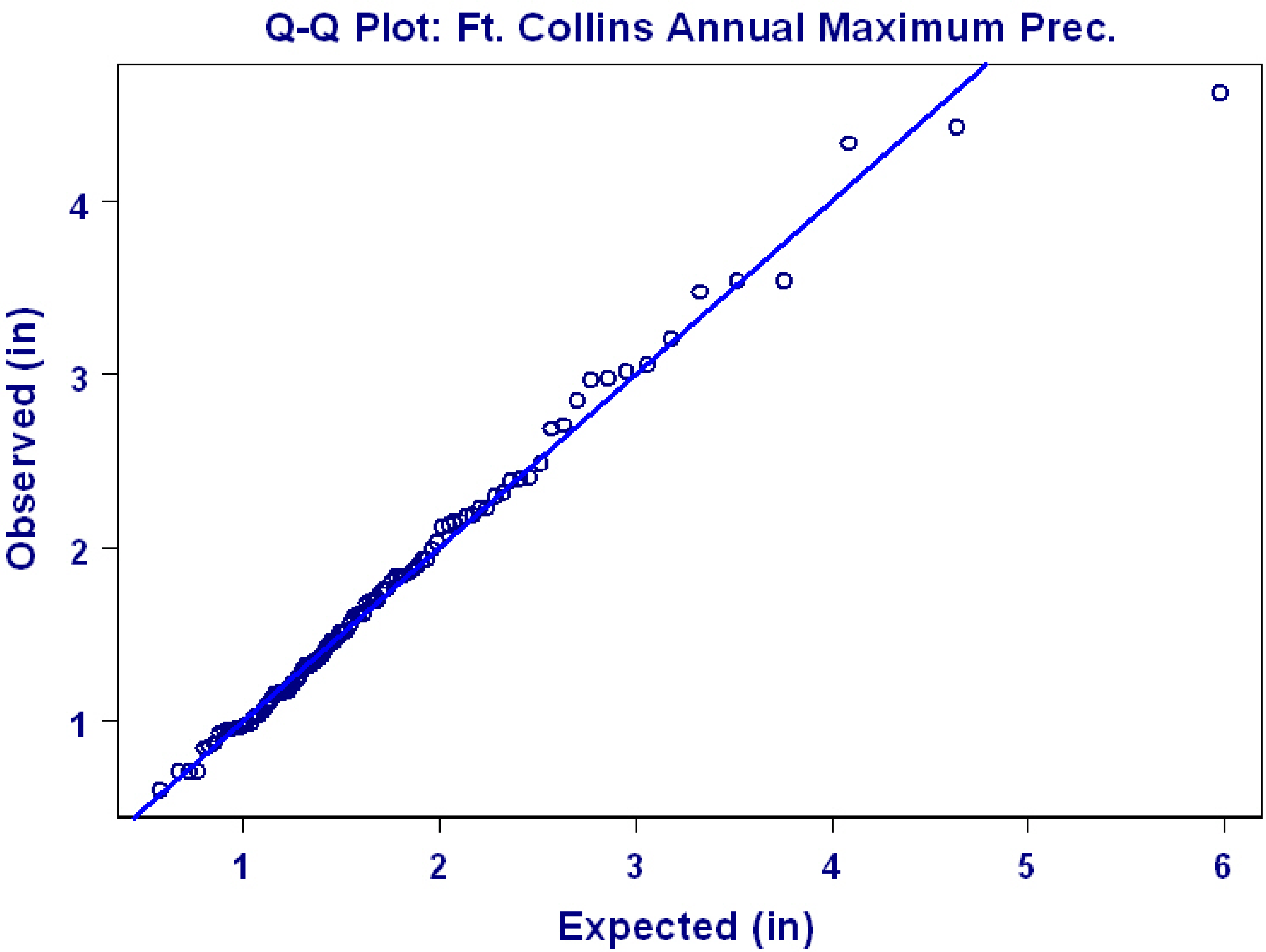
No obvious long-term trend

Recent flood, 28 July 1997

(Damaged campus of Colorado State Univ.)


Fort Collins annual maximum daily precipitation
Precipitation (in)
0
1
2
3
4
1900
1920
1940
1960
1980
2000
Year


Fort Collins annual maximum daily precipitation
Observed relative frequency
Fitted GEV distribution
Probability density function
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
Precipitation (in)


Q-Q Plot: Ft. Collins Annual Maximum Prec.
Observed (in)
Expected (in)
1
2
3
4
5
6

- **Parameter estimates and standard errors**

Parameter	**Estimate**	**(Std. Error)**
Location μ	**1.347**	**(0.617)**
Scale σ	**0.533**	**(0.488)**
Shape ξ	**0.174**	**(0.092)**

-- Likelihood ratio test for $\xi = 0$ (P-value $\approx$ 0.038)

-- 95% confidence interval for shape parameter ξ (based on profile likelihood)

$$0.009 < \xi < 0.369$$

(5) Tails of Distributions

- **Analogue to max stability**

-- X random variable

$Y = X - u$ "excess" over high threshold u, conditional on $X > u$

Then Y has an approximate *generalized Pareto* (GP) distribution for large u:

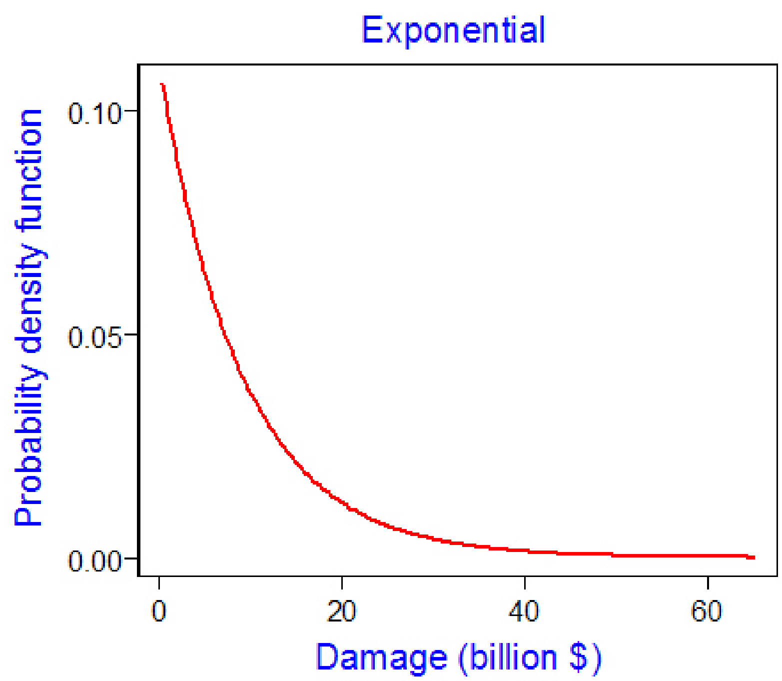
$$\Pr\{Y \le y\} \approx 1 - \left[1 + \xi\,(y/\sigma^*)\right]^{-1/\xi}, \quad y > 0,\ 1 + \xi\,(y/\sigma^*) > 0$$

$\sigma^* > 0$ scale parameter (depends on threshold u)

ξ shape parameter (same interpretation as for GEV dist.)

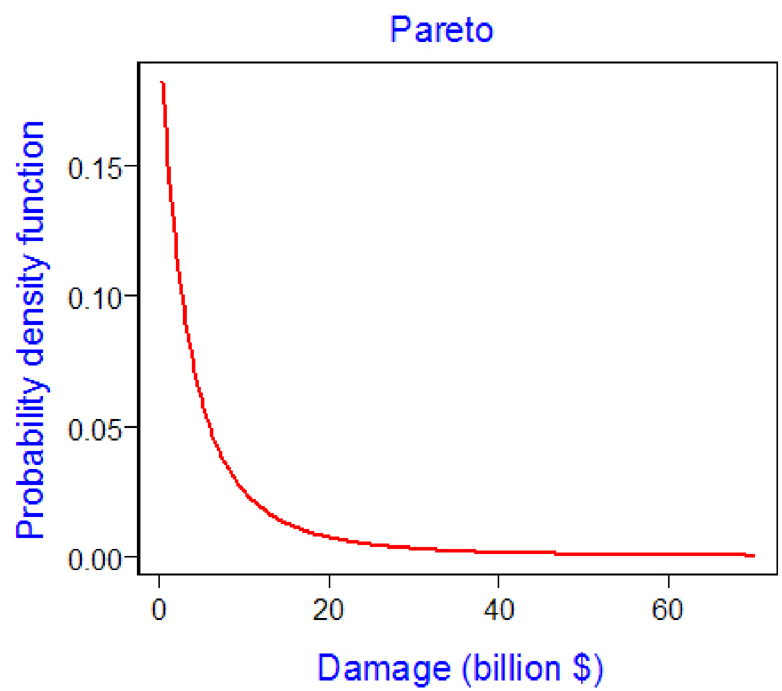
(i) $\xi = 0$ (*exponential type*)

"Light" tail



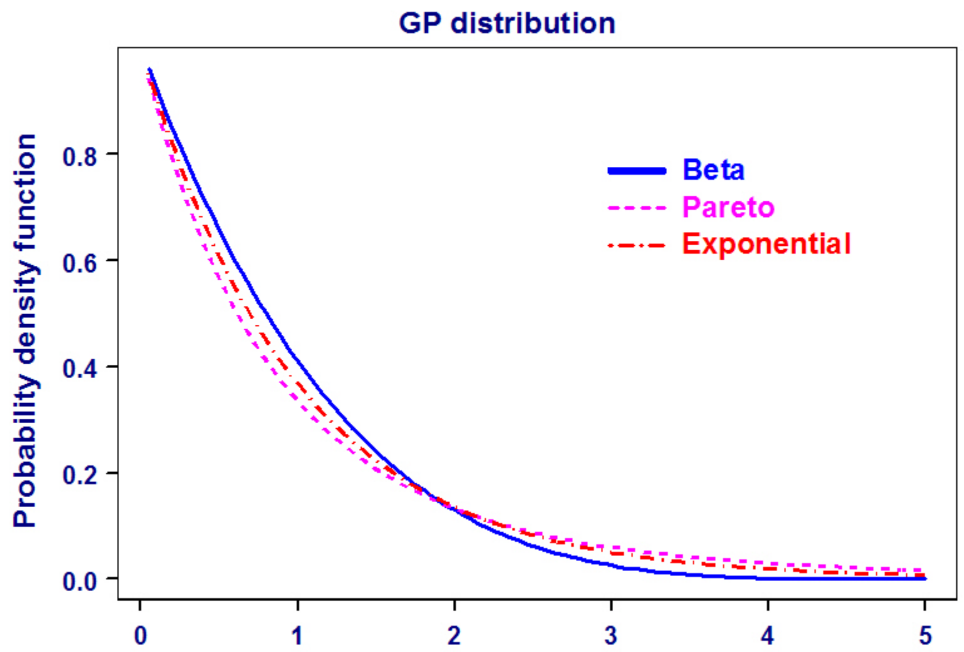
(ii) $\xi > 0$ (*Pareto type*)

"Heavy" tail



(iii) $\xi < 0$ (*beta type*)

Bounded tail [$y < \sigma / (-\xi)$ or $x < u + \sigma / (-\xi)$]



- **Connection between GP & GEV**

-- Maximum $M_T \leq u$ if none of X_t 's exceeds u, $t = 1, 2, \ldots, T$

-- Assume X_t 's independent with common distribution function F

-- Number of exceedances has binomial distribution with

parameters: No. of trials = T

Prob. of "success" = $1 - F(u)$

-- Using Poisson approximation to binomial

$$\Pr\{M_T \leq u\} \approx \exp\{-T[1 - F(u)]\}$$

for large T & u such that $T[1 - F(u)] \approx$ constant

- **Scaling / power laws**

-- "Memoryless" property of exponential distribution (parameter σ^*)

$$\Pr\{Y > y + y' \mid Y > y'\} = \Pr\{Y > y\} = e^{-y/\sigma^*}$$

Suppose *Y* represents life span & has exponential distribution:

Conditional distribution of future survival remains exponential with same scale parameter (no matter how long individual has already survived)

Good model for lifetimes of small birds
(die due to accidents rather than "aging")

-- GP distribution

If condition on higher threshold, then retain same shape parameter but need to rescale (i. e., no longer memoryless)

Suppose excess Y over threshold u has an exact GP distribution with parameters ξ & $\sigma^*(u)$

Then the excess over a higher threshold $u' > u$ has GP distribution with parameters ξ & $\sigma^*(u')$

$$\sigma^*(u') = \sigma^*(u) + \xi(u' - u)$$

- **Hurricane damage**

-- Adjusted data (Remove trends in societal vulnerability), 1925-1995

-- Excess over threshold of u = 6 billion US$ (for year 1995)

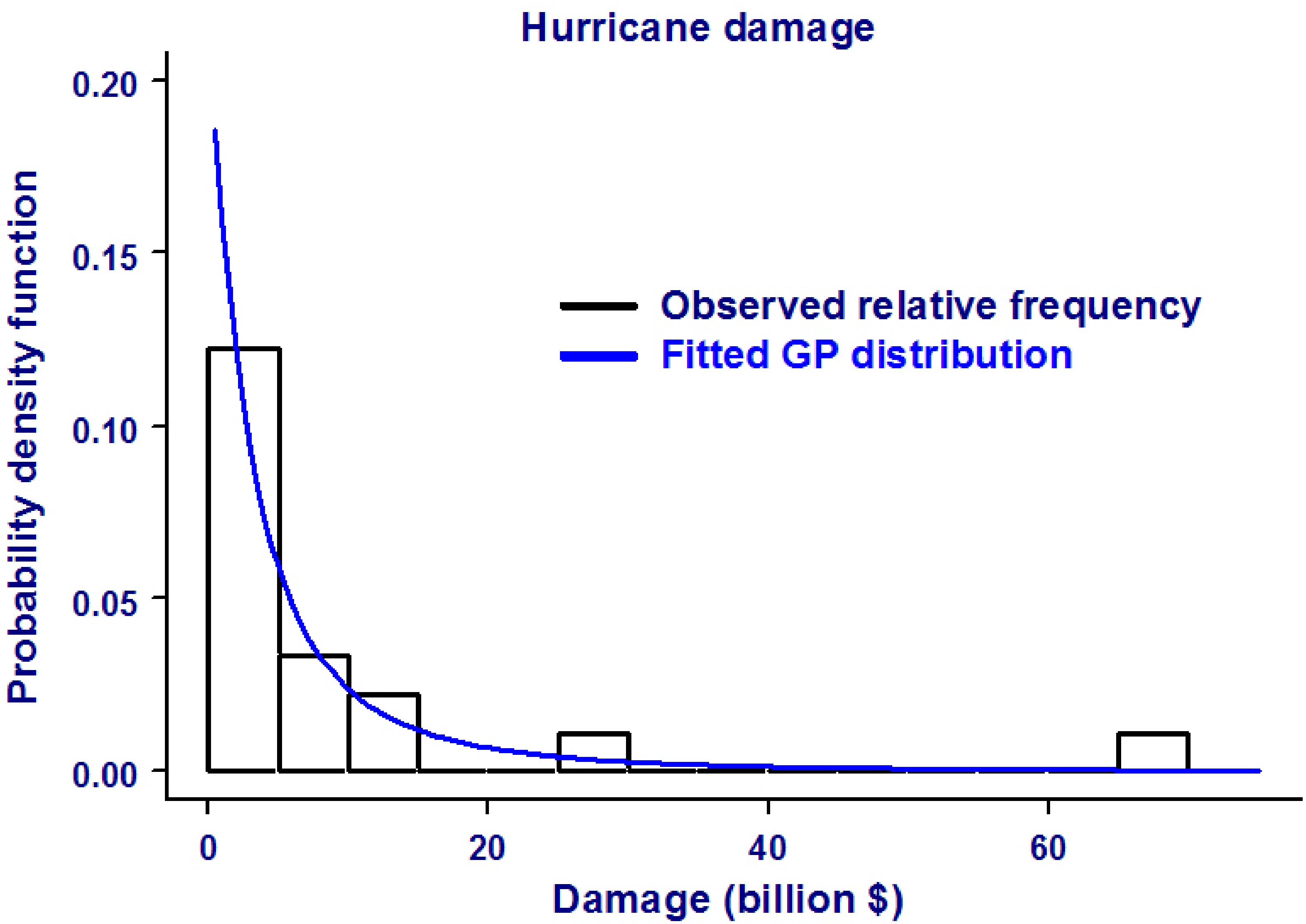
- **Parameter estimates and standard errors**

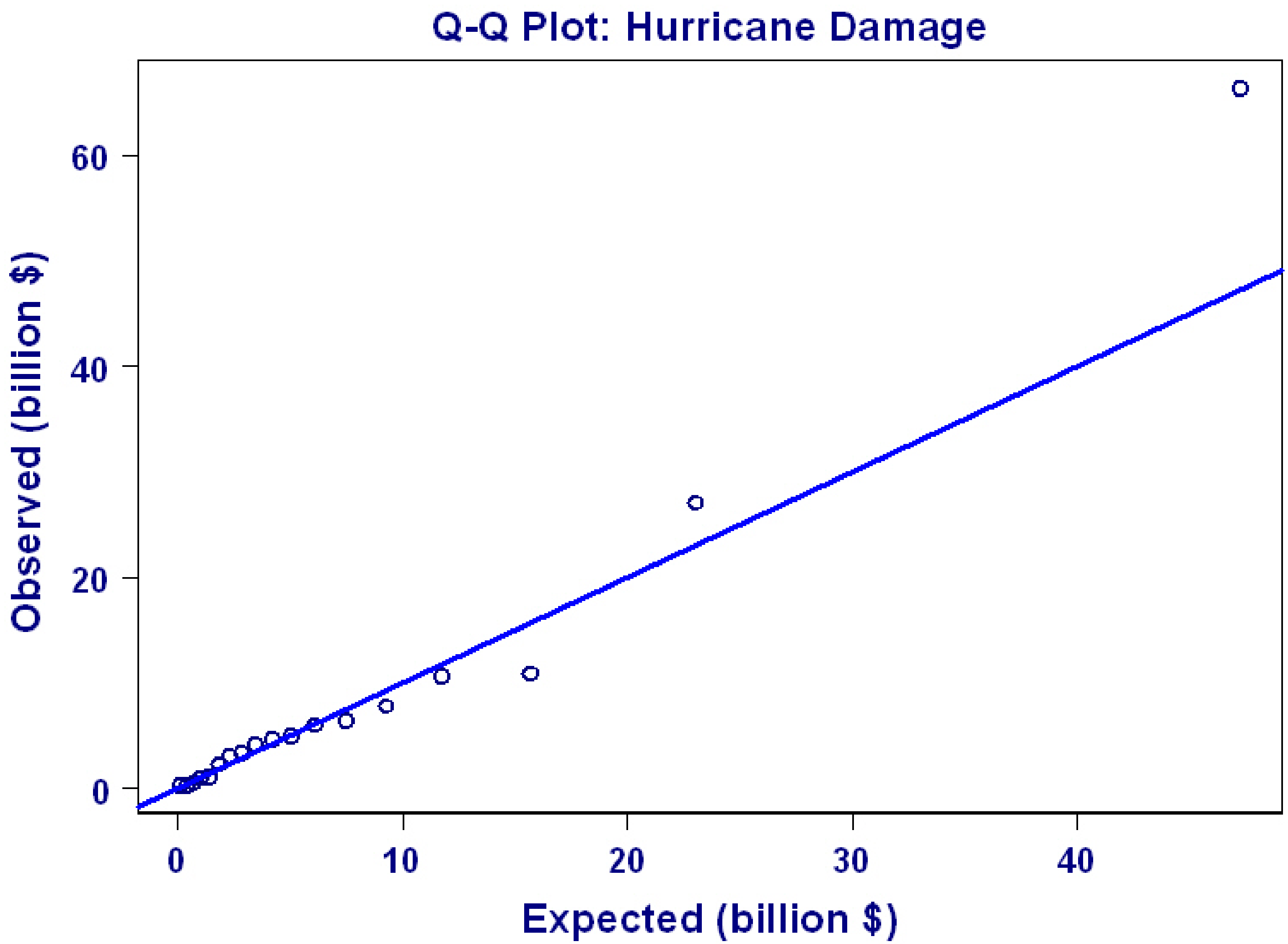
Parameter	Estimate	(Std. Error)
Scale σ^*	4.589	1.817
Shape ξ	0.512	0.341

-- Likelihood ratio test for $\xi = 0$ (P-value $\approx$ 0.018)

-- 95% confidence interval for shape parameter ξ (based on profile likelihood):

$$0.059 < \xi < 1.569$$


Hurricane damage
Probability density function
0.20
0.15
0.10
0.05
0.00
Observed relative frequency
Fitted GP distribution
0
20
40
60
Damage (billion $)


Q-Q Plot: Hurricane Damage
Observed (billion $)
60
40
20
0
0
10
20
30
40
Expected (billion $)

(6) Choice of Threshold

- **Invariance of GP above threshold**

-- Same shape parameter ξ

-- Reparameterize scale parameter: $\sigma^*(\mathrm{adj}) = \sigma^*(u) - \xi\, u$, as u varies

-- Check for stability in parameter estimates as vary threshold

- **Trade-off**

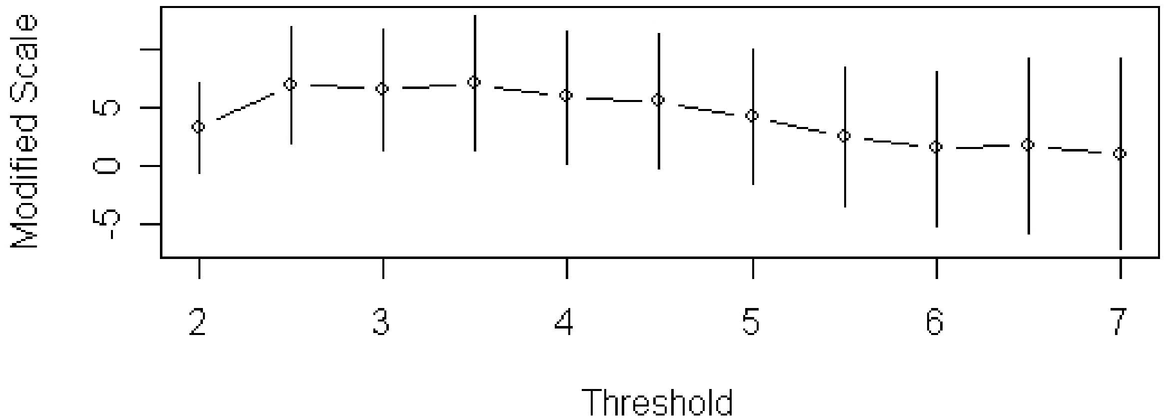
-- Better GP approximation for higher threshold

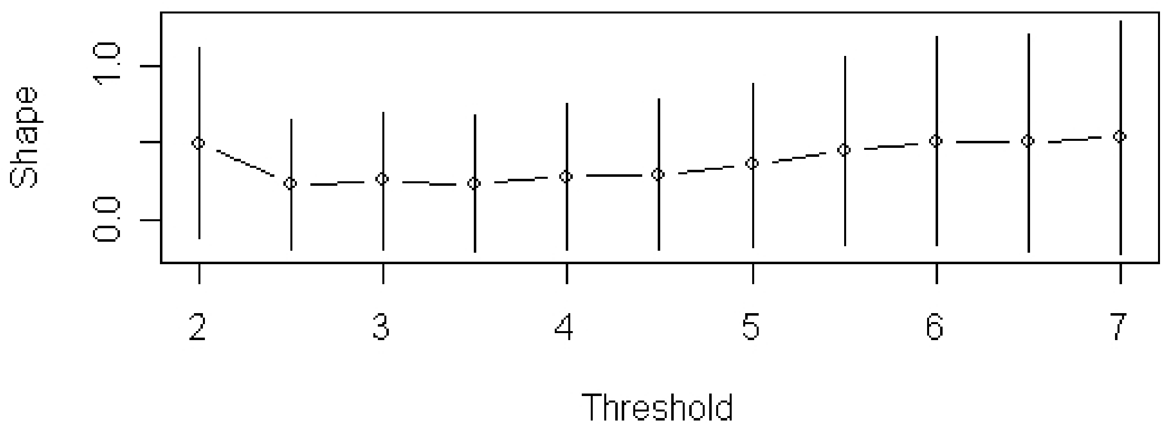
-- More reliable estimation for lower threshold

-- Lack of automatic procedure

- **Hurricane damage example**

-- Vary threshold u from 2 to 7 billion $


Modified Scale
-5
0
5
2
3
4
5
6
7
Threshold


Shape
0.0
1.0
2
3
4
5
6
7
Threshold

(7) De-Clustering

- **Time series models [Autoregressive (AR), short memory]**

-- Lack of clustering at high levels

- **Apparent clustering at high levels**

-- Daily minimum & maximum temperature (strong evidence)

-- Daily precipitation (weak evidence)

- **Lack of automatic procedure for de-clustering**

- **De-clustering procedures**

-- Runs de-clustering

Clusters separated by at least *r* consecutive observations below threshold (*r* = 1, 2, . . .)

Model cluster maxima (instead of individual cluster members)

GEV & GP approximations still valid

(Need to adjust location & scale parameters)

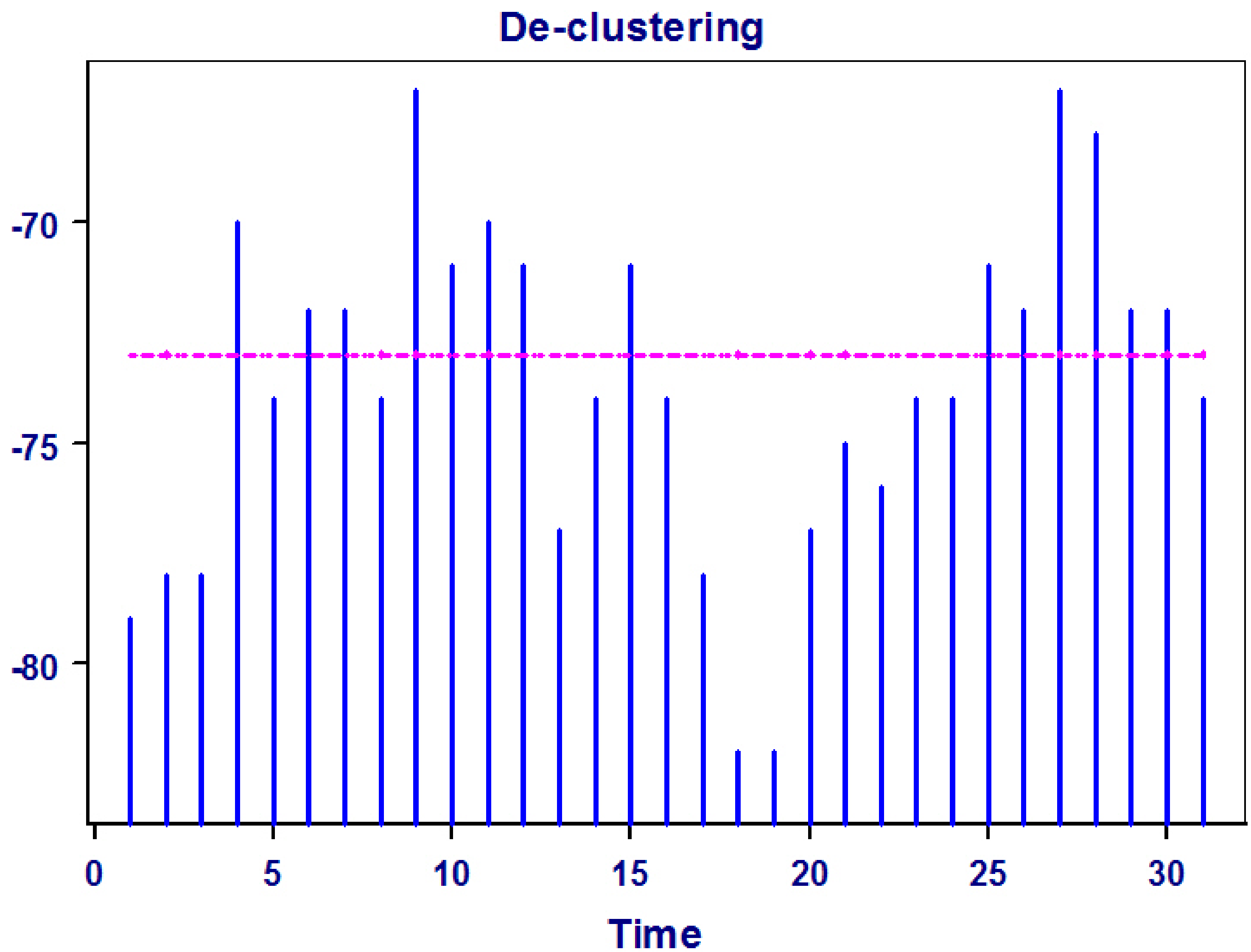
Adjustment depends on extremal index θ, $0 < \theta \leq 1$

θ can be interpreted as inverse of “limiting” mean cluster size

$\theta = 1$ corresponds to negligible dependence at high levels

(e. g., AR process)

Dependence increases as θ decreases


De-clustering
-70
-75
-80
0
5
10
15
20
25
30
Time

- **Phoenix minimum temperature**

-- Phoenix, AZ, USA

Time series of daily minimum temperature (°F) for July-August, 1948-1990

Urban heat island (Warming trend as cities grow)

Phoenix known to have experienced heat island-induced temperature trends

-- Lower tail vs. upper tail

Model lower tail as upper tail after negation

So consider $X^* = -X$, where X denotes daily minimum temperature

- **Fit GP distribution to de-clustered data (ignore any trend for now)**

-- Threshold $u = -73$ °F

-- Use runs de-clustering (separated by $r = 1$ day)

De-clustering	Parameter	Estimate	(Std. Error)
None	Scale σ^*	3.915	(0.303)
$r = 1$		4.167	(0.501)
None	Shape ξ	−0.246	(0.049)
$r = 1$		−0.242	(0.079)

-- Mean cluster size

262 / 115 ≈ 2.3 days (very crude estimate of θ ≈ 0.44)

- **Alternatives to de-clustering**

-- Resampling to estimate standard errors
(avoid throwing away information about extremes)

-- Explicit modeling of temporal dependence at high levels
(e. g., Markov model)

-- Revisit issue later when consider spells of extreme weather
(e. g., heat waves)

(8) Point Process / Peaks over Threshold

- **Rationale**

-- Make more use of information available about upper tail

(even if only interested in obtaining estimate for block maxima)

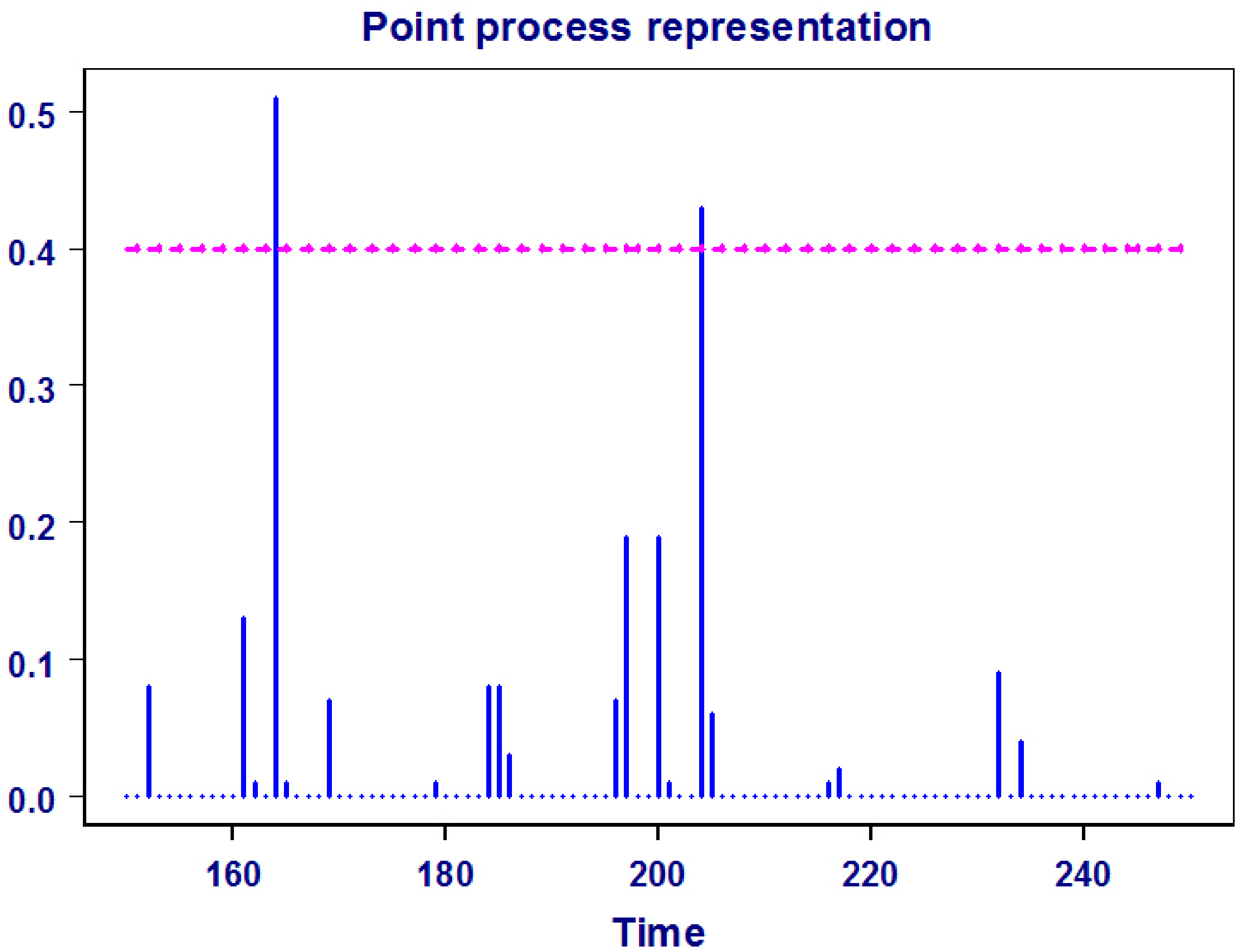
- **Consider process through which extremes arise**

-- *Occurrence*

(e. g., exceedance of high threshold)

-- *Severity*

(e. g., excess over threshold)



Point process representation
0.5
0.4
0.3
0.2
0.1
0.0
160
180
200
220
240
Time

- **Poisson process for rare events**

-- Event is exceedance of high threshold (i. e., $X > u$)

Rate parameter $\lambda > 0$

Pr{no events in [0, T]} = $e^{-\lambda T}$

Mean number of events in [0, T] = λT

- **GP distribution for excess over threshold**

-- Excess $Y = X - u$ (parameters ξ & σ^*)

- **Point process representation**

-- Occurrence & severity of extreme events

(two-dimensional, non-homogenous Poisson process)

-- GEV parameterization

Can relate parameters of GEV (μ, σ, ξ) to parameters of point process (λ, σ^*, ξ):

(i) Shape parameter ξ identical

(ii) $\log \lambda = -(1/\xi) \log[1 + \xi(u - \mu)/\sigma]$

(iii) $\sigma^* = \sigma + \xi(u - \mu)$

Additional detail:

Time scaling constant h

(e.g., for annual maximum of daily data, $h \approx 1/365.25$)

Change of time scale h for GEV(μ, σ, ξ) to h' ($\delta = h/h'$):

$$\sigma' = \sigma\, \delta^{\xi}, \quad \mu' = \mu + [\sigma'(1 - \delta^{-\xi})] / \xi$$

- **Two different approaches to parameter estimation**

-- Orthogonal approach (estimate for two components separately)

Just fit Poisson & GP components separately

Convenient for estimation

Difficult to interpret in presence of covariates

-- GEV re-parameterization (fit both dimensions simultaneously)

More difficult to estimate

But interpretable even with covariates

- **Fort Collins daily precipitation**

-- Now analyze daily data instead of just annual maxima (but ignore annual cycle for now)

Estimation Method	Parameter	Estimate
(i) Orthogonal	Rate λ	10.6 per yr.
(u = 0.395 in)	Scale σ^*	0.322
	Shape ξ	0.212
(ii) Point process	Location μ	1.383
(u = 0.395 in,	Scale σ	0.532
h = 1/365.25)	Shape ξ	0.212

(9) Risk Communication (Under Stationarity)

- **Assume Stationarity (for now)**

-- Unchanging climate

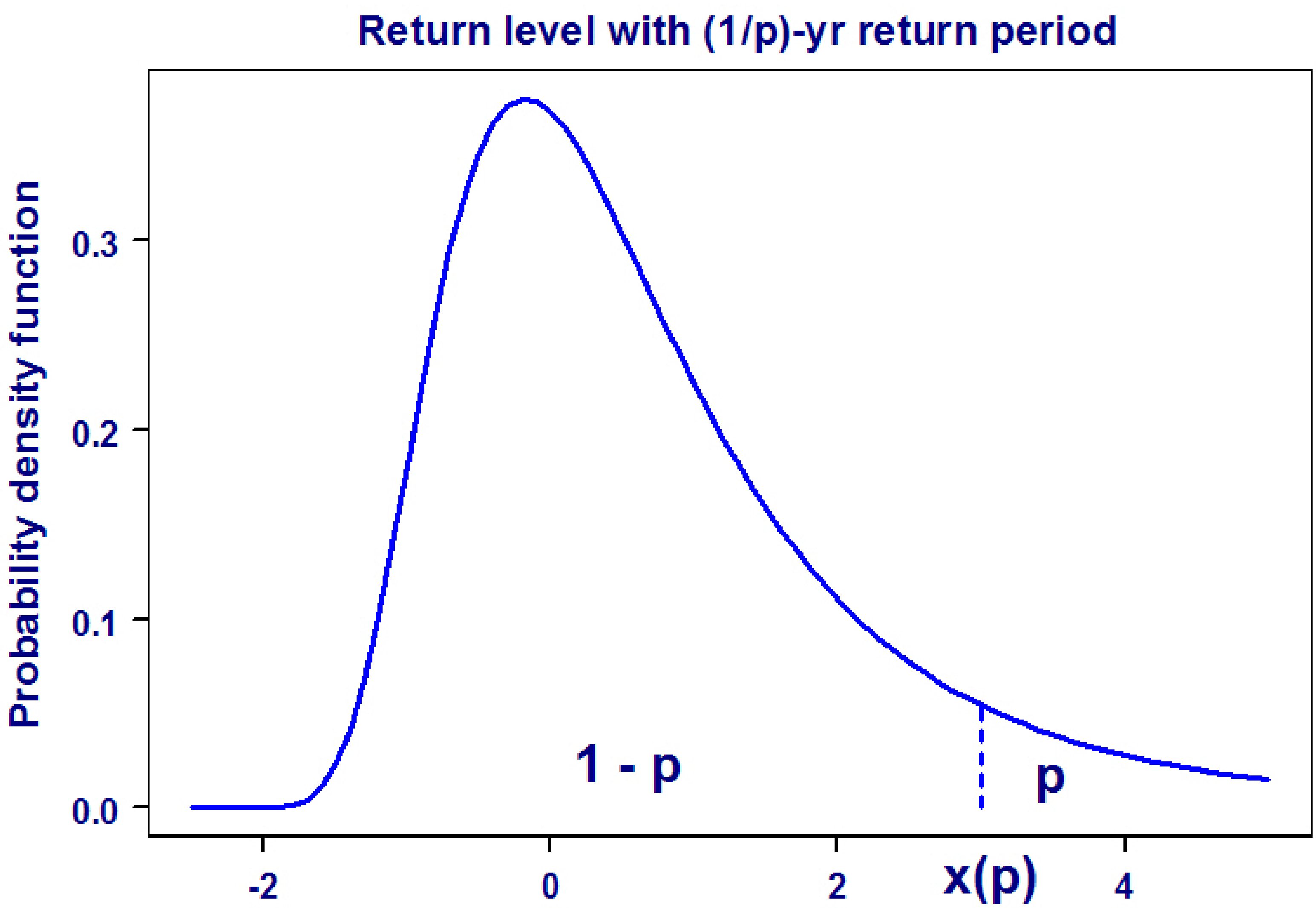
- **Return period / return level**

-- Return level with (1/*p*)-yr return period

$$x(p) = F^{-1}(1 - p)$$

Quantile of cumulative distribution function *F* (e. g., GEV)

(e. g., *p* = 0.01 corresponds to 100-yr return period)


Return level with (1/p)-yr return period
Probability density function
0.0
0.1
0.2
0.3
-2
0
2
4
x(p)
1 - p
p

- **GEV distribution**

$$x(p) = \mu - (\sigma/\xi)\ \{1 - [-\log(1 - p)]\}^{-\xi}$$

Confidence interval: Re-parameterize replacing location parameter μ with $x(p)$ & use profile likelihood

-- Fort Collins precipitation example (annual maxima)

Estimated 100-yr return period: 5.10 in

95% confidence interval based on profile likelihood:

$$3.93 \text{ in} < x(p) < 8.00 \text{ in}$$

-- Similar approach for GP distribution

(except complication to take into account rate of occurrence of exceedances of threshold)