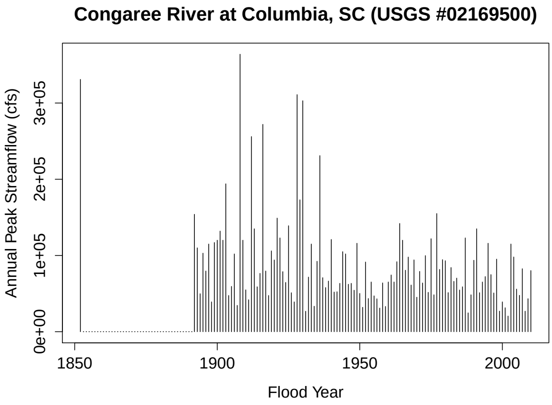
Historic Floods and 1% Chance Flood Estimation

Philip Yates

Department of Mathematics
Saint Michael's College
pyates@smcvt.edu

March 6, 2012

Congaree River: Columbia, SC



Congaree River: Columbia, SC

Flood Frequency Data

Annual Peak Flows

The maximum momentary peak discharge in each year of record

- The year of record begins the previous October and ends in September of the current year
- Peak flows are measured in cubic feet per second (cfs)

Flood Frequency Data

Annual Peak Flows

The maximum momentary peak discharge in each year of record

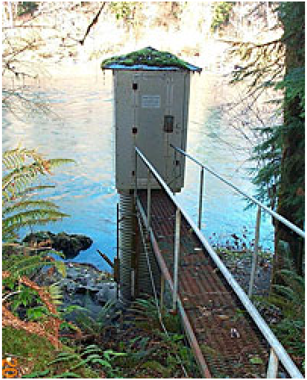
- The year of record begins the previous October and ends in September of the current year
- Peak flows are measured in cubic feet per second (cfs)

Historical Flood Data

Any observation of flood stage or conditions made before actual flood data were collected systematically

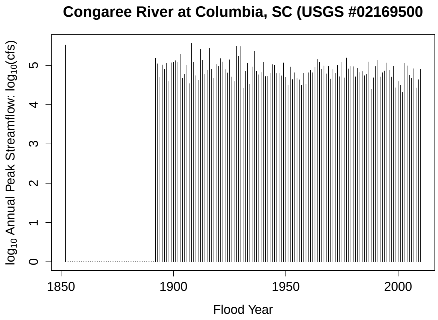
- Data collected from old newspapers, diaries, museums, libraries, etc.

River Gage

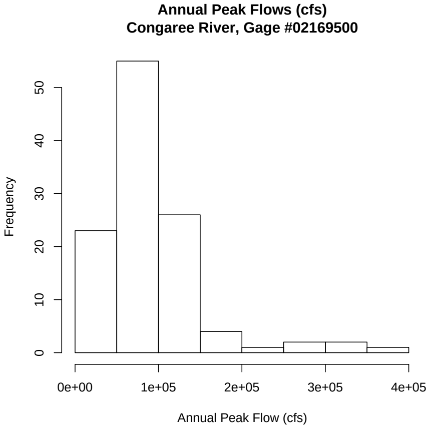


Gage #12041200: Hoh River at U.S. Highway 101 near Forks, WA

Congaree River, Gage #02169500



Congaree River, Gage #02169500



1865: Columbia, SC

- Ruins, as seen from the State House, 1865.
- General Sherman's Union troops were slowed entering Columbia by a major flood on the Congaree River.

Modeling Flood Frequency Data

Annual peak flows are assumed to be independent and identically distributed.

Let $Y = \log_{10} X$, where X is the annual peak flow.

$$f_Y(y) = \frac{(y-\gamma)^{\alpha-1}\exp[-(y-\gamma)/\beta]}{\beta^\alpha \Gamma(\alpha)},$$

$\alpha > 0, \beta > 0, y > \gamma$, where:

α — the shape parameter

β — the scale parameter

γ — the shift or location parameter

Modeling Flood Frequency Data

Bulletin #17B (1981)

- FEMA uses it as the trade standard

Modeling Flood Frequency Data

Bulletin #17B (1981)

- FEMA uses it as the trade standard
- Method of moments used to estimate parameters of log-Pearson type III distribution

Modeling Flood Frequency Data

Bulletin #17B (1981)

- FEMA uses it as the trade standard
- Method of moments used to estimate parameters of log-Pearson type III distribution
- When dealing with mixed populations, splits data into number of groups and fits a separate curve for each group

Modeling Flood Frequency Data

Bulletin #17B (1981)

- FEMA uses it as the trade standard
- Method of moments used to estimate parameters of log-Pearson type III distribution
- When dealing with mixed populations, splits data into number of groups and fits a separate curve for each group
- When collecting annual peak flow data, estimates the 99th percentile of the log-Pearson type III distribution; this is used as an estimate for the 1 percent chance — FEMA uses this in regulatory policy for floodplains

Other Works

American River flood frequency analyses (1999)

- Used both systematic annual peak flows and one historical flood observation from 1862

Other Works

American River flood frequency analyses (1999)

- Used both systematic annual peak flows and one historical flood observation from 1862
- Used EMA, the Expected Moments Algorithm (Cohn, et al., 1997) to fit the data to a log-Pearson type III distribution

Other Works

American River flood frequency analyses (1999)

- Used both systematic annual peak flows and one historical flood observation from 1862
- Used EMA, the Expected Moments Algorithm (Cohn, et al., 1997) to fit the data to a log-Pearson type III distribution
- Recommended the independent identically distributed approach to flood estimation

Other Works

American River flood frequency analyses (1999)

- Used both systematic annual peak flows and one historical flood observation from 1862
- Used EMA, the Expected Moments Algorithm (Cohn, et al., 1997) to fit the data to a log-Pearson type III distribution
- Recommended the independent identically distributed approach to flood estimation
- Tentative use of mixed models as remedy for certain types of non-stationarity in the flood data

Other Works (Mixtures)

Singh (1987a,b) used a mixture of two normal distributions and two lognormal distributions in order to consider an observed flood that is composed of two component distributions

Other Works (Mixtures)

Singh (1987a,b) used a mixture of two normal distributions and two lognormal distributions in order to consider an observed flood that is composed of two component distributions

Hirschboeck (1987) & Diehl and Potter (1987) discussed various reasons to believe that there may be two or more components to the observed flood's distribution

Other Works (Mixtures)

Singh (1987a,b) used a mixture of two normal distributions and two lognormal distributions in order to consider an observed flood that is composed of two component distributions

Hirschboeck (1987) & Diehl and Potter (1987) discussed various reasons to believe that there may be two or more components to the observed flood's distribution

- in the southeastern region of US, tropical storms

Other Works (Mixtures)

Singh (1987a,b) used a mixture of two normal distributions and two lognormal distributions in order to consider an observed flood that is composed of two component distributions

Hirschboeck (1987) & Diehl and Potter (1987) discussed various reasons to believe that there may be two or more components to the observed flood's distribution

- in the southeastern region of US, tropical storms
- in the western, and even midwestern and northeastern, regions of US, snowmelt

Planned Work

- Investigate methods to find parameter estimates for a mixture of normals

Planned Work

- Investigate methods to find parameter estimates for a mixture of normals
- Develop an estimate of the standard error for the 99th percentile of a finite mixture model with two components

Planned Work

- Investigate methods to find parameter estimates for a mixture of normals
- Develop an estimate of the standard error for the 99th percentile of a finite mixture model with two components
- Use these methods on the Congaree flood data in order to find the 1 percent chance flood estimate (the 99th percentile of the fitted distribution)

Planned Work

- Investigate methods to find parameter estimates for a mixture of normals
- Develop an estimate of the standard error for the 99th percentile of a finite mixture model with two components
- Use these methods on the Congaree flood data in order to find the 1 percent chance flood estimate (the 99th percentile of the fitted distribution)
- Incorporate the 1852 flood into the flood frequency analysis

Planned Work

- Investigate methods to find parameter estimates for a mixture of normals
- Develop an estimate of the standard error for the 99th percentile of a finite mixture model with two components
- Use these methods on the Congaree flood data in order to find the 1 percent chance flood estimate (the 99th percentile of the fitted distribution)
- Incorporate the 1852 flood into the flood frequency analysis
- Future work and research projects

Finite Mixture Model

Suppose Y is a random variable or vector that takes values from sample space $\mathcal{Y}$.

$$p(y) = \pi_1 f_1(y) + \ldots + \pi_k f_k(y) \quad (y \in \mathcal{Y}),$$

where

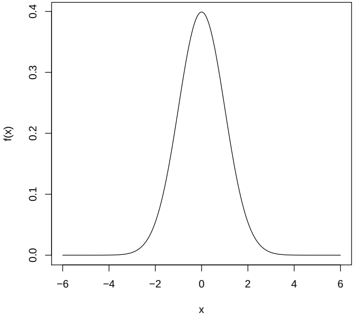
$$\pi_j > 0, \quad j = 1, \ldots, k; \quad \pi_1 + \ldots + \pi_k = 1$$

and

$$f_j(\cdot) \geq 0, \quad \int_{\mathcal{Y}} f_j(y)\,\mathrm{d}x = 1, \quad j = 1, \ldots, k,$$

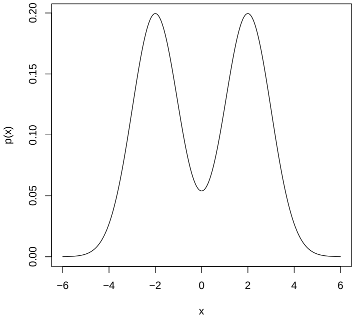
then Y has a finite mixture distribution

Finite Mixture Model



- Standard Normal p.d.f.: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

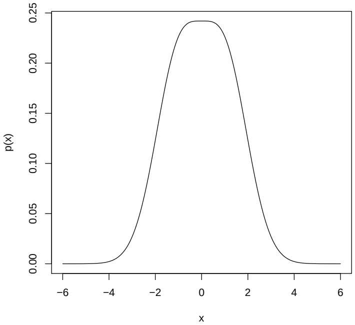
Finite Mixture Model



- Finite Mixture Distribution:

$$p(x) = 0.5 * \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+2)^2} + 0.5 * \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}$$

Finite Mixture Model



- Finite Mixture Distribution:

$$p(x) = 0.5 * \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x+1)^2} + 0.5 * \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}$$

Finite Mixture Model – Basic Concepts

Likelihood Function

If the set of random variables $Y_1, Y_2, \ldots, Y_n$ denotes a random sample from a population with probability density (or mass) function $f(y|\theta)$, then

$$\begin{aligned} L(\theta|y_1, y_2, \ldots, y_n) &= f(y_1, y_2, \ldots, y_n|\theta) \\ &= f(y_1|\theta) \times f(y_2|\theta) \times \cdots \times f(y_n|\theta), \end{aligned}$$

is the likelihood function. This will typically be denoted as $L(\theta)$.

Finite Mixture Model

Likelihood Function for Finite Mixture Model

$$L(\psi) = \prod_{i=1}^{n} p(y_i|\psi) = \prod_{i=1}^{n} \left[\sum_{j=1}^{k} \pi_j f(y_i|\boldsymbol{\theta}_j) \right]$$

$\pi_j f(y_i|\boldsymbol{\theta}_j)$ component keeps track of y_i's contribution to entire mixture density

If it was known which component of the mixture the observation was from, the data would be fully categorized:

Complete Data Likelihood

$$\{x_i, i = 1, \ldots, n\} = \{(y_i, \mathbf{z}_i); i = 1, \ldots, n\},$$

where

$\mathbf{z}_i = (z_{ij} = 1, \ldots, k)$ — indicator vector of length k with 1 in the position corresponding to the appropriate component and zeros elsewhere

$\mathbf{x}$ — complete data (not observed directly)

$\mathbf{y}$ — incomplete data ($\mathbf{x}$ is indirectly observed via $\mathbf{y}$)

Complete Data Likelihood

Complete Data Likelihood

$$g(x_1, \ldots, x_n | \psi) = \prod_{i=1}^{n} \prod_{j=1}^{k} \pi_j^{z_{ij}} f_j(y_i | \theta_j)^{z_{ij}}$$

Complete Data Likelihood

Complete Data Likelihood

$$g(x_1, \ldots, x_n|\psi) = \prod_{i=1}^{n}\prod_{j=1}^{k} \pi_j^{z_{ij}} f_j(y_i|\theta_j)^{z_{ij}}$$

Complete Data Log-likelihood

$$l(\psi) = \sum_{i=1}^{n} \mathbf{z}_i^{\mathbf{T}} \mathbf{V}(\boldsymbol{\pi}) + \sum_{i=1}^{n} \mathbf{z}_i^{\mathbf{T}} \mathbf{U}_i(\boldsymbol{\theta}),$$

where $\log \pi_j$ is the j^{th} component of $\mathbf{V}(\boldsymbol{\pi})$ and $\log f_j(y_i|\boldsymbol{\theta}_j)$ is the j^{th} component of $\mathbf{U}_i(\boldsymbol{\theta})$

Observable Log-Likelihood

Define for each i and j

$$\mathbf{w}_i = w_{ij} = E(\mathbf{z}_i | y_i, \psi') = \frac{\pi_j' f_j(y_i|\boldsymbol{\theta}_j')}{p(y_i|\psi')}$$

Observable Log-Likelihood

Define for each i and j

$$\mathbf{w}_i = w_{ij} = E(\mathbf{z}_i | y_i, \psi') = \frac{\pi'_j f_j(y_i|\boldsymbol{\theta}'_j)}{p(y_i|\psi')}$$

Observable Log-Likelihood Function

$$Q(\psi|\psi') = E(\log g(\mathbf{x}|\psi)|\mathbf{y}, \psi') =$$
$$\sum_{i=1}^{n} \mathbf{w}_i^{\mathbf{T}} \mathbf{V}(\pi) + \sum_{i=1}^{n} \mathbf{w}_i^{\mathbf{T}} \mathbf{U}_i(\theta)$$

Maximum Likelihood Estimation

Example: Suppose $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson distribution with mean λ.

$$f(y_i|\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad i = 1, \ldots, n$$

Maximum Likelihood Estimation

Example: Suppose $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson distribution with mean λ.

$$
\begin{aligned}
f(y_i|\lambda) &= \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad i = 1, \ldots, n \\
L(\lambda) &= \frac{\lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!}
\end{aligned}
$$

Maximum Likelihood Estimation

Example: Suppose $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson distribution with mean λ.

$$
\begin{aligned}
f(y_i|\lambda) &= \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad i = 1, \ldots, n \\
L(\lambda) &= \frac{\lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!} \\
\log L(\lambda) &= (\sum_{i=1}^{n} y_i) \log \lambda - n\lambda - \log(\prod_{i=1}^{n} y_i!)
\end{aligned}
$$

Maximum Likelihood Estimation

Example: Suppose $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson distribution with mean λ.

$$\begin{aligned}
f(y_i|\lambda) &= \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad i = 1, \ldots, n \\
L(\lambda) &= \frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{\prod_{i=1}^n y_i!} \\
\log L(\lambda) &= (\sum_{i=1}^n y_i) \log \lambda - n\lambda - \log(\prod_{i=1}^n y_i!) \\
\frac{\partial \log L(\lambda)}{\partial \lambda} &= \frac{\sum_{i=1}^n y_i}{\lambda} - n
\end{aligned}$$

Maximum Likelihood Estimation

Example: Suppose $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson distribution with mean λ.

$$
\begin{aligned}
f(y_i|\lambda) &= \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad i = 1, \ldots, n \\
L(\lambda) &= \frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{\prod_{i=1}^n y_i!} \\
\log L(\lambda) &= (\sum_{i=1}^n y_i) \log \lambda - n\lambda - \log(\prod_{i=1}^n y_i!) \\
\frac{\partial \log L(\lambda)}{\partial \lambda} &= \frac{\sum_{i=1}^n y_i}{\lambda} - n \\
\hat{\lambda} &= \frac{\sum_{i=1}^n y_i}{n} = \overline{y}
\end{aligned}
$$

Mixture of Two Normals

$$f(y_i|\mu_0, \sigma_0^2, \mu_1, \sigma_1^2, \pi) = (1-\pi)\phi\left(\frac{y_i - \mu_0}{\sqrt{\sigma_0^2}}\right) + \pi\phi\left(\frac{y_i - \mu_1}{\sqrt{\sigma_1^2}}\right),$$

To implement ECM (Meng & Rubin, 1993) algorithm, let

$$w_i = \frac{\pi\phi\left(\frac{y_i-\mu_1}{\sqrt{\sigma_1^2}}\right)}{(1-\pi)\phi\left(\frac{y_i-\mu_0}{\sqrt{\sigma_0^2}}\right) + \pi\phi\left(\frac{y_i-\mu_1}{\sqrt{\sigma_1^2}}\right)}.$$

Mixture of Two Normals

E-step

$$Q(\psi|\psi^{(p)}) = \sum_{i=1}^{n} w_i^{(p)} \log(\pi) +$$

$$(1 - w_i^{(p)}) \log(1-\pi) + w_i^{(p)} \log\phi\left(\frac{y_i - \mu_1}{\sqrt{\sigma_1^2}}\right) +$$

$$(1 - w_i^{(p)}) \log\phi\left(\frac{y_i - \mu_0}{\sqrt{\sigma_0^2}}\right).$$

Mixture of Two Normals

CM-steps

$$\pi^{(p+1)} = \frac{1}{n}\sum_{i=1}^{n} w_i^{(p)}$$

$$\mu_0^{(p+1)} = \frac{\sum_{i=1}^{n}(1-w_i^{(p)})y_i}{\sum_{i=1}^{n}(1-w_i^{(p)})}$$

$$\mu_1^{(p+1)} = \frac{\sum_{i=1}^{n} w_i^{(p)} y_i}{\sum_{i=1}^{n} w_i^{(p)}}$$

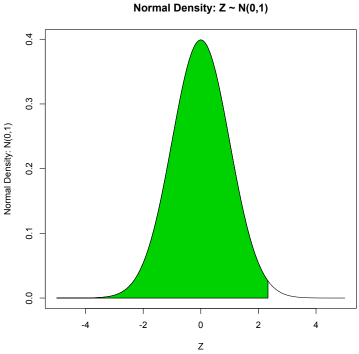
Mixture of Two Normals

CM-steps

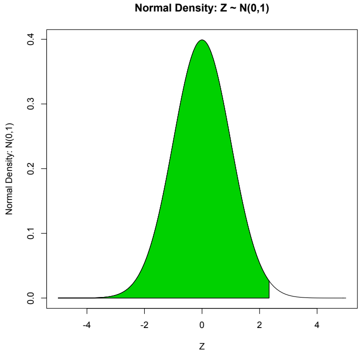
$$\sigma_0^{2(p+1)} = \frac{\sum_{i=1}^{n}(1 - w_i^{(p)})(y_i - \mu_0^{(p+1)})^2}{\sum_{i=1}^{n}(1 - w_i^{(p)})}$$

$$\sigma_1^{2(p+1)} = \frac{\sum_{i=1}^{n} w_i^{(p)}(y_i - \mu_1^{(p+1)})^2}{\sum_{i=1}^{n} w_i^{(p)}}$$

1% Chance Flood Estimation



1% Chance Flood Estimation



The 99th percentile of the standard normal distribution is $\Phi^{-1}(0.99) = 2.326$.

1% Chance Flood Estimation

To find the 1 percent chance flood when the distribution is from a finite mixture model, one needs to find the 99th percentile of the finite mixture distribution. For example, if the distribution was a finite mixture model of normal densities, then the 100-year flood, Q, is

$$\int_{-\infty}^{\log_{10} Q} (1-\pi) f_0(y|\mu_0, \sigma_0^2) + \pi f_1(y|\mu_1, \sigma_1^2) dy = 0.99,$$

where $f_j(y) \sim N(\mu_j, \sigma_j^2)$ and $\log_{10} Q$ is the logged value of the 1 percent chance flood.

To find $\log_{10} \hat{Q}$, first obtain the 99th percentile from each component of the mixing distribution.

1% Chance Flood Estimation

These percentiles from each mixing component are used as end points to search for the root of

$$\int_{-\infty}^{\log_{10}\hat{Q}} (1-\pi) f_0(y|\mu_0, \sigma_0^2) + \pi f_1(y|\mu_1, \sigma_1^2) dy - 0.99 = 0.$$

In order to obtain a $100(1-\alpha)\%$ confidence interval for $\log_{10} Q$ in this example, a delta method argument can be made to find the standard error of $\log_{10} Q$

1% Chance Flood Estimation – Basic Terms

Example: Total Derivative of a Function

The total derivative of $f(t, x, y)$ with respect to t is

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial t}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

1% Chance Flood Estimation – Basic Terms

Example: Total Derivative of a Function

The total derivative of $f(t, x, y)$ with respect to t is

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial t}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Delta Method

For some sequence of random variables X_n satisfying

$$\sqrt{n}[X_n - \theta] \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{\mathcal{D}} N(0, \sigma^2[g'(\theta)]^2).$$

1% Chance Flood Estimation

Define the quantile function for a finite mixture model with two normal densities as:

$$\begin{aligned}\Gamma(\boldsymbol{\theta}, x(\boldsymbol{\theta}, 0.99)) &= \int_{-\infty}^{x} \left\{ \tau \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(y-\mu_1)^2} + (1-\tau)\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2}(y-\mu_0)^2} \right\} dy \\ &= 0.99,\end{aligned}$$

with the chain rule yielding

$$\nabla_\tau x(\boldsymbol{\theta}, 0.99) = -\frac{\nabla_\tau \Gamma(\boldsymbol{\theta}, x)|_{x=x(\boldsymbol{\theta},0.99)}}{\nabla_x \Gamma(\boldsymbol{\theta}, x)|_{x=x(\boldsymbol{\theta},0.99)}} \quad \text{Uryasev (2000),}$$

where $\boldsymbol{\theta}^{\mathbf{T}} = (\mu_0, \sigma_0^2, \mu_1, \sigma_1^2, \tau)^{\mathbf{T}}$.

1% Chance Flood Estimation

The numerator of $\nabla_\tau x(\boldsymbol{\theta}, 0.99)$ is,

$$\begin{aligned}
\nabla_\tau \Gamma(\boldsymbol{\theta}, x)|_{x=x(\boldsymbol{\theta},0.99)} &= \int_{-\infty}^{x} \left\{ \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(y-\mu_1)^2} - \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2}(y-\mu_0)^2} \right\} dy \\
&= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(y-\mu_1)^2} dy - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2}(y-\mu_0)^2} dy \\
&= \Phi\left(\frac{x(\boldsymbol{\theta}, 0.99) - \mu_1}{\sigma_1} \right) - \Phi\left(\frac{x(\boldsymbol{\theta}, 0.99) - \mu_0}{\sigma_0} \right),
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal probability distribution.

1% Chance Flood Estimation

The numerator for the gradient of this quantile function with respect to μ_0 and μ_1 is

$$\begin{aligned}
\nabla_{\mu_0}\Gamma(\boldsymbol{\theta}, x)\big|_{x=x(\boldsymbol{\theta},0.99)} &= \int_{-\infty}^{x} \frac{(1-\tau)(y-\mu_0)}{\sqrt{2\pi}\sigma_0^3} e^{-\frac{1}{2\sigma_0^2}(y-\mu_0)^2} dy \\
\nabla_{\mu_1}\Gamma(\boldsymbol{\theta}, x)\big|_{x=x(\boldsymbol{\theta},0.99)} &= \int_{-\infty}^{x} \frac{\tau(y-\mu_1)}{\sqrt{2\pi}\sigma_1^3} e^{-\frac{1}{2\sigma_1^2}(y-\mu_1)^2} dy
\end{aligned}$$

1% Chance Flood Estimation

The numerator for the gradient of this quantile funciton with respect to σ_0^2 and σ_1^2 is

$$\begin{aligned}
\nabla_{\sigma_0^2}\Gamma(\boldsymbol{\theta},x)|_{x=x(\boldsymbol{\theta})} &= \int_{-\infty}^{x}\left\{\frac{1-\tau}{2\sqrt{2\pi}\sigma_0^3}e^{-\frac{1}{2\sigma_0^2}(y-\mu_0)^2}\left(\frac{(y-\mu_0)^2}{\sigma_0^2}-1\right)\right\}dy \\
\nabla_{\sigma_1^2}\Gamma(\boldsymbol{\theta},x)|_{x=x(\boldsymbol{\theta})} &= \int_{-\infty}^{x}\left\{\frac{\tau}{2\sqrt{2\pi}\sigma_1^3}e^{-\frac{1}{2\sigma_1^2}(y-\mu_1)^2}\left(\frac{(y-\mu_1)^2}{\sigma_1^2}-1\right)\right\}dy.
\end{aligned}$$

1% Chance Flood Estimation

The denominator for the gradient of this quantile function with respect to any of the parameters in $\boldsymbol{\theta}$ is,

$$
\begin{aligned}
\nabla_x \Gamma(\boldsymbol{\theta}, x)\big|_{x=x(\boldsymbol{\theta},0.99)} &= p(x(\tau, 0.99)) \\
&= \tau \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x(\boldsymbol{\theta},0.99)-\mu_1)^2} \\
&\quad +(1-\tau)\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2}(x(\boldsymbol{\theta},0.99)-\mu_0)^2}.
\end{aligned}
$$

1% Chance Flood Estimation

Confidence Interval

$$\text{Estimate} \pm \text{Confidence Level} \times \text{Standard Error}$$

1% Chance Flood Estimation

Confidence Interval

$$\text{Estimate} \pm \text{Confidence Level} \times \text{Standard Error}$$

$100(1-\alpha)\%$ Confidence Interval for $\log_{10} Q$

$$\log_{10} \hat{Q} \pm z_{\frac{\alpha}{2}} \left(\nabla \log_{10} \hat{Q}^{\mathbf{T}} I_Y(\hat{\psi})^{-1} \nabla \log_{10} \hat{Q} \right)^{1/2},$$

where I_Y is the observed information matrix.

Observed Information Matrix

Observed Information Matrix

$$\mathbf{I}_Y(\psi) = E(\mathbf{B}(\mathbf{x}|\psi)|\mathbf{y},\psi) \\ -E(\mathbf{S}(\mathbf{x}|\psi)\mathbf{S}(\mathbf{x}|\psi)^{\mathsf{T}}|\mathbf{y},\psi) + \mathbf{S}^*(\mathbf{y}|\psi)\mathbf{S}^*(\mathbf{y}|\psi)^{\mathsf{T}}.$$

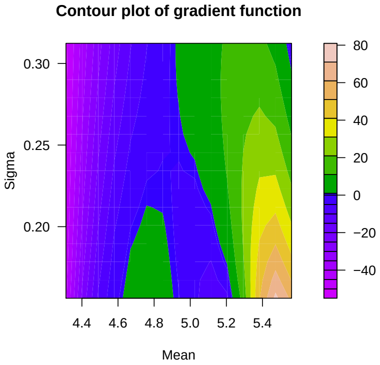
Observed Information Matrix

Observed Information Matrix

$$\mathbf{I}_Y(\psi) = E(\mathbf{B}(\mathbf{x}|\psi)|\mathbf{y}, \psi) \\ -E(\mathbf{S}(\mathbf{x}|\psi)\mathbf{S}(\mathbf{x}|\psi)^{\mathbf{T}}|\mathbf{y}, \psi) + \mathbf{S}^*(\mathbf{y}|\psi)\mathbf{S}^*(\mathbf{y}|\psi)^{\mathbf{T}}.$$

This is the conditional expected complete data information matrix minus the expected information matrix for the conditional distribution of the complete data, $\mathbf{x}$, given the incomplete data, $\mathbf{y}$, and ψ.

Congaree River



The results here suggest two mixing components are needed for the $\log_{10}$ of the Congaree River's annual peak flows

Congaree River

Mixture of Normals $\hat{\mu_0} = 4.855215$, $\hat{\sigma_0} = 0.213704$, $\hat{\mu_1} = 5.4717257$, $\hat{\sigma_1} = 0.0583816$, $\hat{\pi} = 0.0385812$

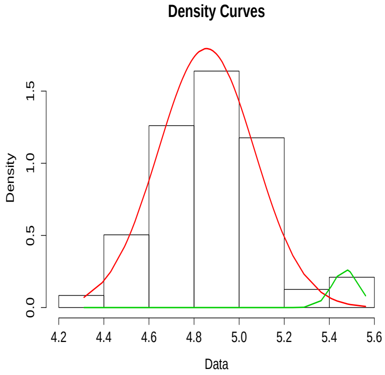
Point Estimate $\log_{10} \hat{Q} = 5.51417$; $\hat{Q} = 326,716$ cfs

Standard Error Estimate 0.0354481

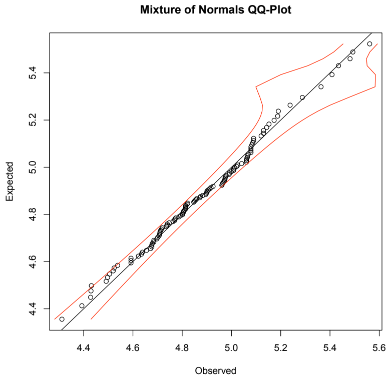
95% CI Q: 278,415 cfs to 383,396 cfs

Number of Flood Years: 119

Congaree River



Congaree River



Congaree River: 1852, 1892 to 2010

What information do we actually have?

Congaree River: 1852, 1892 to 2010

What information do we actually have?

- 1852 – approximately 331,000 cfs

Congaree River: 1852, 1892 to 2010

What information do we actually have?

- 1852 – approximately 331,000 cfs
- 1853 to 1891 – unobserved

Congaree River: 1852, 1892 to 2010

What information do we actually have?

- 1852 – approximately 331,000 cfs
- 1853 to 1891 – unobserved
  - Safe to assume these annual peak flows are less than 331,000 cfs

Congaree River: 1852, 1892 to 2010

What information do we actually have?

- 1852 – approximately 331,000 cfs
- 1853 to 1891 – unobserved
    - Safe to assume these annual peak flows are less than 331,000 cfs
    - Treat these "censored" observations as if they were from a truncated normal probability distribution:

$$f(x|\mu,\sigma,331000) = \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}}{\int_{-\infty}^{331,000}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx}$$

- 1892 to 2010 – observed from flood gauge

Congaree River: 1852, 1892 to 2010

The mixture distribution with censored observations is:

$$p(y_i|\psi) = \sum_{j=1}^{k} \pi_j \left\{f_j(y_i|\boldsymbol{\theta}_j)\right\}^{\delta_i} \left\{F_j(y_i|\boldsymbol{\theta}_j)\right\}^{1-\delta_i},$$

where $\delta_i = 1$ if y_i is observed and 0 if y_i is censored.
In this analysis:

$$\begin{aligned} f_j(y_i) &= \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{1}{2\sigma_j^2}(y_i-\mu_j)^2\right] \\ F_j(y_i) &= \int_{-\infty}^{331{,}000} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{1}{2\sigma_j^2}(t-\mu_j)^2\right] dt \end{aligned}$$

Congaree River: 1852, 1892 to 2010

Calculate the posterior probabilities of category membership for the *i*th observation, conditional on y_i and $\boldsymbol{\psi}'$:

$$w_{ij} = E\left(z_{ij} | y_i, \boldsymbol{\psi}'\right) = \frac{\pi'_j f_j(y_i|\boldsymbol{\theta}'_j)^{\delta_i} F_j(y_i|\boldsymbol{\theta}'_j)^{1-\delta_i}}{p(y_i|\boldsymbol{\psi}')}$$

Then:

1. Calculate the observed log-likelihood function.

Congaree River: 1852, 1892 to 2010

Calculate the posterior probabilities of category membership for the ith observation, conditional on y_i and $\boldsymbol{\psi}'$:

$$w_{ij} = E\left(z_{ij} | y_i, \boldsymbol{\psi}'\right) = \frac{\pi_j' f_j(y_i|\boldsymbol{\theta}_j')^{\delta_i} F_j(y_i|\boldsymbol{\theta}_j')^{1-\delta_i}}{p(y_i|\boldsymbol{\psi}')}$$

Then:

1. Calculate the observed log-likelihood function.
2. Use ECM to find the MLE's of the mixture distribution.

Congaree River: 1852, 1892 to 2010

Calculate the posterior probabilities of category membership for the ith observation, conditional on y_i and $\boldsymbol{\psi}'$:

$$w_{ij} = E\left(z_{ij}|y_i, \boldsymbol{\psi}'\right) = \frac{\pi_j' f_j(y_i|\boldsymbol{\theta}_j')^{\delta_i} F_j(y_i|\boldsymbol{\theta}_j')^{1-\delta_i}}{p(y_i|\boldsymbol{\psi}')}$$

Then:

1. Calculate the observed log-likelihood function.
2. Use ECM to find the MLE's of the mixture distribution.
3. Estimate the 99th quantile of the mixture distribution.

Congaree River: 1852, 1892 to 2010

Calculate the posterior probabilities of category membership for the *i*th observation, conditional on y_i and $\boldsymbol{\psi}'$:

$$w_{ij} = E\left(z_{ij} | y_i, \boldsymbol{\psi}'\right) = \frac{\pi_j' f_j(y_i|\boldsymbol{\theta}_j')^{\delta_i} F_j(y_i|\boldsymbol{\theta}_j')^{1-\delta_i}}{p(y_i|\boldsymbol{\psi}')}$$

Then:

1. Calculate the observed log-likelihood function.
2. Use ECM to find the MLE's of the mixture distribution.
3. Estimate the 99th quantile of the mixture distribution.
4. Use the total derivative of a function and the delta method to estimate the standard error of the 99th percentile of the mixture distribution.

Congaree River: 1852, 1892 to 2010

Mixture of Normals $\hat{\mu_0} = 4.855537$, $\hat{\sigma_0} = 0.2139423$, $\hat{\mu_1} = 5.477557$, $\hat{\sigma_1} = 0.05249058$, $\hat{\pi} = 0.04280238$

Point Estimate $\log_{10} \hat{Q} = 5.519481$; $\hat{Q} = 330,736$ cfs

Standard Error Estimate 0.01577529

  95% CI Q: 308,008 cfs to 355,140 cfs

Number of Flood Years: 159

Inclusion of the historic flood of 1852 increased the point estimate of the 1% flood by a minimal amount (about 4000 cfs).

The inclusion of the historic flood and the censored years increased the number of observations to 159, thus shrinking the interval estimate of the 1% chance flood.

Recommendations

Use the gradient contour plot developed by Lindsay (1983) to decide whether or not additional components need to be added to a finite mixing distribution

If the plot indicates that additional components are needed, then fit the $\log_{10}$ of the annual peak flood flows to a mixture of two normal distributions

If additional components are not needed, proceed with Bulletin #17B in order to obtain that point estimate and confidence interval for the annual peak flood flows

Future Research

How would short-term dependence affect the estimation of the 1 percent chance flood? The 1929 (March 1, 1929) and 1930 (October 3, 1929) flood years could pose a problem for the Congaree River.

Pre-dam analysis versus post-dam analysis: Dreher Shoals Dam (a.k.a. Lake Murray Dam) built in 1930. Built on the Saluda River, it lies about approximately 10 miles west of where the Broad River and Saluda River meet to form the Congaree River in Columbia.

Future Collaboration – Undergraduate

Possible Undergraduate Research Projects:

- Flood frequency analysis on the Winooski River – Two gauges still in use
  - Gage #04286000 – Montpelier, VT (1912 – 2010)
  - Gage #04290500 – near Essex Junction, VT (1928 – 2009)
- Fit a finite mixture of Gumbel distributions and find the standard error for the 1% chance flood
- Fit a finite mixture model where the components are from two different distributions?
  - Gamma and Normal?