Combining p-values

Patrick Rubin-Delanchy
patrick.rubin-delanchy@bristol.ac.uk

School of Mathematics, University of Bristol

November 22, 2018

Meta-analysis

- Meta-analysis: "Examination of data from a number of independent studies of the same subject, in order to determine overall trends" (Oxford dictionary).

Meta-analysis

- Meta-analysis: "Examination of data from a number of independent studies of the same subject, in order to determine overall trends" (Oxford dictionary).
- Since statistical studies often amount to reporting a p-value for a certain effect, meta-analysis is in many contexts taken to be synonymous with combining p-values.

Meta-analysis

- Meta-analysis: "Examination of data from a number of independent studies of the same subject, in order to determine overall trends" (Oxford dictionary).
- Since statistical studies often amount to reporting a p-value for a certain effect, meta-analysis is in many contexts taken to be synonymous with combining p-values.
- We will consider the formal framework:

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0, 1],$$

$$\text{versus,}$$

$$H_1 : P_i \overset{ind}{\sim} F_i.$$

Meta-analysis

- Meta-analysis: "Examination of data from a number of independent studies of the same subject, in order to determine overall trends" (Oxford dictionary).
- Since statistical studies often amount to reporting a p-value for a certain effect, meta-analysis is in many contexts taken to be synonymous with combining p-values.
- We will consider the formal framework:

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0, 1],$$

versus,

$$H_1 : P_i \overset{ind}{\sim} F_i.$$

- Note that we are assuming independence between the p-values under both hypotheses. We'll touch on dependence later.

Anomaly detection

- In many monitoring applications, you will hear people talk about generating hundreds/thousands/millions of "anomaly scores" and combining them using advanced techniques. What's usually going on is that the anomaly scores are p-values, and these are being combined in more or less intricate ways, in fact often by the very simple techniques we are about to learn.

Anomaly detection

- In many monitoring applications, you will hear people talk about generating hundreds/thousands/millions of "anomaly scores" and combining them using advanced techniques. What's usually going on is that the anomaly scores are p-values, and these are being combined in more or less intricate ways, in fact often by the very simple techniques we are about to learn.
- The relevance of this hypothesis testing framework is therefore obvious, and the applications are important. For example, "There has been substantial recent growth in the demand for surveillance of public health and clinical indicators. In the USA this has been primarily motivated by anxiety over bioterrorism following the anthrax attacks in 2001 (Bravataet al., 2004) although any threats to human health may be included: for example, the Biosense Program of the US Center for Disease Control and Prevention [...] is a major initiative in 'syn-dromic surveillance', where the aim is rapid identification of clustered outbreaks of disease. In contrast, in the UK the driving motivation behind surveillance of clinical care arises from 'scandals' such as the Bristol heart babies (Spiegelhalteret al., 2002) and the Shipman murder case (Aylinet al., 2003), and the public prominence given to hospital-acquired infections such as MRSA and *Clostridium difficile*." (Spiegelhalter, 2012)

Fisher's method

"When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance. It is sometimes desired, taking account only of these probabilities, and not of the detailed composition of the data from which they are derived, which may be of very different kinds, to obtain a single test of the significance of the aggregate, based on the product of the probabilities individually observed." (Fisher, 1932)

- What Fisher is suggesting is to use the product $\prod_{i=1}^{N} P_i$ as a measure of global significance.
- The product is significant when small, so to stay consistent with our convention (and many other reasons) we will instead consider:

$$T = -2 \sum \log(P_i),$$

which is equivalent, and gives a test that rejects for large values of T.

Distribution of $-2\log(\text{uniform})$

A Gamma(α, β) distribution has density:

$$f_\Gamma(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, e^{-\beta x}\,.$$

Let U be a uniform random variable on $[0, 1]$. Then

$$X = -2\log(U) \sim \text{Gamma}(\,\textvisiblespace\,,\,\textvisiblespace\,),$$

since,

Distribution of $-2\log(\text{uniform})$

A Gamma(α, β) distribution has density:

$$f_\Gamma(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, e^{-\beta x}\,.$$

Let U be a uniform random variable on $[0, 1]$. Then

$$X = -2\log(U) \sim \text{Gamma}(\text{␣}\,,\text{␣}\,),$$

since,

$$\mathbb{P}(X \le x) =$$

Distribution of $-2\log(\text{uniform})$

A Gamma(α, β) distribution has density:

$$f_\Gamma(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, e^{-\beta x}\, .$$

Let U be a uniform random variable on $[0, 1]$. Then

$$X = -2\log(U) \sim \text{Gamma}(\;\_\;,\;\_\;),$$

since,

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \mathbb{P}(-2\log(U) \leq x), \\
&= \mathbb{P}(U \geq e^{-x/2}), \\
&= 1 - e^{-x/2},
\end{aligned}$$

so that

$$f_X(x) = \frac{1}{2}\, e^{-x/2} = f_\Gamma(x;\;\_\;,\;\_\;).$$

Distribution of $-2\log(\text{uniform})$

A $\text{Gamma}(\alpha, \beta)$ distribution has density:

$$f_\Gamma(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \, e^{-\beta x} \, .$$

Let U be a uniform random variable on $[0, 1]$. Then

$$X = -2\log(U) \sim \text{Gamma}(1, 1/2),$$

since,

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(-2\log(U) \leq x), \\ &= \mathbb{P}(U \geq e^{-x/2}), \\ &= 1 - e^{-x/2}, \end{aligned}$$

so that

$$f_X(x) = \frac{1}{2} e^{-x/2} = f_\Gamma(x; 1, 1/2).$$

Distribution of $-2\sum \log(\text{uniform})$

- For integer k, the Gamma$(k/2, 1/2)$ distribution is known as a chi-square distribution with k degrees of freedom, denoted χ^2_k, with distribution function $F_{\chi^2}(x, k)$.

Distribution of $-2\sum \log(\text{uniform})$

- For integer k, the Gamma$(k/2, 1/2)$ distribution is known as a chi-square distribution with k degrees of freedom, denoted χ^2_k, with distribution function $F_{\chi^2}(x, k)$.
- It is a well-known fact (and straightforward to prove) that the sum of independent chi-square variables with respective degrees of freedom $k_1, \ldots, k_n$ is itself chi-square, with $k_1 + \cdots + k_n$ degrees of freedom.

Distribution of $-2\sum\log(\text{uniform})$

- For integer k, the Gamma$(k/2, 1/2)$ distribution is known as a chi-square distribution with k degrees of freedom, denoted χ^2_k, with distribution function $F_{\chi^2}(x, k)$.

直观：卡方分布是正太分布生成的

- It is a well-known fact (and straightforward to prove) that the sum of independent chi-square variables with respective degrees of freedom $k_1, \ldots, k_n$ is itself chi-square, with $k_1 + \cdots + k_n$ degrees of freedom.
- Consider n independent uniform random variables $U_1, \ldots, U_n$ on $[0, 1]$. Since $-2\log(U_i) \sim \text{Gamma}(1, 1/2)$, we have:

$$-2\sum_{i=1}^{n}\log(U_i) \sim \chi^2_{\text{␣}}.$$

Distribution of $-2\sum \log(\text{uniform})$

- For integer k, the Gamma$(k/2, 1/2)$ distribution is known as a chi-square distribution with k degrees of freedom, denoted χ^2_k, with distribution function $F_{\chi^2}(x, k)$.
- It is a well-known fact (and straightforward to prove) that the sum of independent chi-square variables with respective degrees of freedom $k_1, \ldots, k_n$ is itself chi-square, with $k_1 + \cdots + k_n$ degrees of freedom.
- Consider n independent uniform random variables $U_1, \ldots, U_n$ on $[0, 1]$. Since $-2\log(U_i) \sim$ Gamma$(1, 1/2)$, we have:

$$-2\sum_{i=1}^{n} \log(U_i) \sim \chi^2_{2n}.$$

- It therefore follows that under the null hypothesis,

$$T = -2\sum_{i=1}^{n} \log(P_i) \sim \chi^2_{2n}.$$

A bit of intuition

Fisher's method allows a majority (although perhaps not overwhelming) of tests to be inconclusive while still detecting an effect if there are a few, strong signals. For example:

- with ten p-values respectively equal to 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.01, 0.01, 0.01, we obtain $-2\sum \log(p_i) \approx 37$, $1 - F_{\chi^2}(37, 20) \approx 0.01$.
- with ten p-values equal to 0.353 (the average of the above), we obtain $-2\sum \log(p_i) \approx 21$, $1 - F_{\chi^2}(21, 20) \approx 0.41$

A bit of intuition

Fisher's method allows a majority (although perhaps not overwhelming) of tests to be inconclusive while still detecting an effect if there are a few, strong signals. For example:

- with ten p-values respectively equal to 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.01, 0.01, 0.01, we obtain $-2\sum \log(p_i) \approx 37$, $1 - F_{\chi^2}(37, 20) \approx 0.01$.
- with ten p-values equal to 0.353 (the average of the above), we obtain $-2\sum \log(p_i) \approx 21$, $1 - F_{\chi^2}(21, 20) \approx 0.41$ (it's a consequence of the inequality of arithmetic and geometric means).

A specified alternative

The Beta(α, β) distribution has density:

$$f_{\text{Beta}}(x) = \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $\alpha, \beta > 0$ and

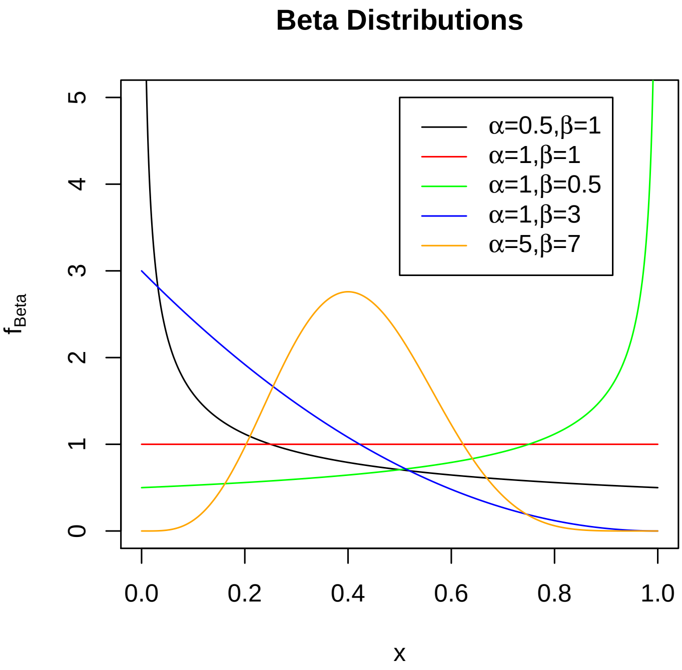
$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Now, consider the hypothesis testing framework:

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0, 1],$$

versus,

$$H_1 : P_i \overset{i.i.d.}{\sim} \text{Beta}(\alpha, 1), \quad \alpha < 1.$$



Beta Distributions
α=0.5,β=1
α=1,β=1
α=1,β=0.5
α=1,β=3
α=5,β=7
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
f_{Beta}
x

Likelihood ratio

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0,1],$$

$$\text{versus,}$$

$$H_1 : P_i \overset{i.i.d.}{\sim} \text{Beta}(\alpha, 1), \quad \alpha < 1.$$

The likelihood ratio is:

$$T = \frac{\text{density under } H_1}{\text{density under } H_0} = \frac{\text{␣}}{\text{␣}}.$$

Likelihood ratio

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0,1],$$
$$\text{versus,}$$
$$H_1 : P_i \overset{i.i.d.}{\sim} \text{Beta}(\alpha,1), \quad \alpha < 1.$$

The likelihood ratio is:

$$T = \frac{\text{density under } H_1}{\text{density under } H_0} = \frac{\mathrm{B}(\alpha,\beta)^{-n}\left(\prod_{i=1}^{n} P_i\right)^{\alpha-1}}{␣}.$$

这里其实是要检验P，通过T来检验P，而不是之前的实验了，检验实验结果的检验?

Likelihood ratio

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0,1],$$
$$\text{versus,}$$
$$H_1 : P_i \overset{i.i.d.}{\sim} \text{Beta}(\alpha, 1), \quad \alpha < 1.$$

The likelihood ratio is:

$$T = \frac{\text{density under } H_1}{\text{density under } H_0} = \frac{\mathrm{B}(\alpha,\beta)^{-n} \left(\prod_{i=1}^{n} P_i\right)^{\alpha-1}}{1}.$$

Likelihood ratio

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0,1],$$

versus,

$$H_1 : P_i \overset{i.i.d.}{\sim} \text{Beta}(\alpha, 1), \quad \alpha < 1.$$

The likelihood ratio is:

$$T = \frac{\text{density under } H_1}{\text{density under } H_0} = \frac{\mathrm{B}(\alpha,\beta)^{-n} \left(\prod_{i=1}^{n} P_i\right)^{\alpha-1}}{1}.$$

This is equivalent to:

Likelihood ratio

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0,1],$$

versus,

$$H_1 : P_i \overset{i.i.d.}{\sim} \text{Beta}(\alpha, 1), \quad \alpha < 1.$$

The likelihood ratio is:

$$T = \frac{\text{density under } H_1}{\text{density under } H_0} = \frac{\mathrm{B}(\alpha,\beta)^{-n} \left(\prod_{i=1}^{n} P_i\right)^{\alpha-1}}{1}.$$

This is equivalent to:

$$-2\sum_{i=1}^{n} \log(P_i).$$

因为和参数无关

It therefore follows that Fisher's method is the uniformly most powerful combiner against all alternatives of the form

$$H_1 : P_i \overset{i.i.d.}{\sim} \text{Beta}(\alpha, 1), \quad \alpha < 1.$$

所以fisher的方法在这个H1中是最好检验方法使得power最大

A reasonable assumption?

Is there any reason to think a p-value could be Beta$(\alpha, 1), \alpha > 1$ under H_1?

Consider the following example (from Heard, Nicholas A., and Rubin-Delanchy, P. "Choosing between methods of combining-values." *Biometrika* 105.1 (2018): 239-246.)

$$
\begin{aligned}
H_0 &: X \sim \text{exponential}(\lambda_0), \\
&\text{versus,} \\
H_1 &: X \sim \text{exponential}(\lambda_1), \quad \lambda_1 < \lambda_0.
\end{aligned}
$$

这里的X不是P是什么意思

是不是说我假设X服从这些分布，
看看生成的P是什么分布的

Remember that the exponential distribution has density

$$f_{\text{exp}}(x) = \lambda\, \mathrm{e}^{-\lambda x},$$

and distribution function

$$F_{\text{exp}}(x) = 1 - \mathrm{e}^{-\lambda x}\,.$$

A reasonable assumption?

The LRT is

$$\frac{\lambda_1 \, e^{-\lambda_1 X}}{\lambda_0 \, e^{-\lambda_0 X}} = \frac{\lambda_1}{\lambda_0} \, e^{(\lambda_0 - \lambda_1) X},$$

which is equivalent to:

A reasonable assumption?

The LRT is

$$\frac{\lambda_1 \, e^{-\lambda_1 X}}{\lambda_0 \, e^{-\lambda_0 X}} = \frac{\lambda_1}{\lambda_0} \, e^{(\lambda_0 - \lambda_1) X},$$

which is equivalent to:

$$X.$$

这里X是最好的检验方法，power最大

We will therefore use as a test statistic $T = X$ and compute the p-value:

$$p = \mathbb{P}_0(T \geq t) = \mathbb{P}_0(X \geq t) = e^{-\lambda_0 t}.$$

意思其实是这个T服从指数分布，然后就能算?算什么

Now consider the distribution of this p-value under the alternative. We have:

$$\mathbb{P}_1(P \le x) = \mathbb{P}_1(e^{-\lambda_0 T} \le x),$$

是不是说p值就是在假设下发生的概率，
然后我们现在要算这个概率在H1假设下的分布

Now consider the distribution of this p-value under the alternative. We have:

$$\begin{aligned}\mathbb{P}_1(P \leq x) &= \mathbb{P}_1(\mathrm{e}^{-\lambda_0 T} \leq x), \\ &= \mathbb{P}_1\{-\lambda_0 T \leq \log(x)\},\end{aligned}$$

Now consider the distribution of this p-value under the alternative. We have:

$$\begin{aligned}
\mathbb{P}_1(P \leq x) &= \mathbb{P}_1(\mathrm{e}^{-\lambda_0 T} \leq x), \\
&= \mathbb{P}_1\{-\lambda_0 T \leq \log(x)\}, \\
&= \mathbb{P}_1\left\{T \geq -\frac{\log(x)}{\lambda_0}\right\},
\end{aligned}$$

Now consider the distribution of this p-value under the alternative. We have:

$$
\begin{aligned}
\mathbb{P}_1(P \leq x) &= \mathbb{P}_1(\mathrm{e}^{-\lambda_0 T} \leq x), \\
&= \mathbb{P}_1\{-\lambda_0 T \leq \log(x)\}, \\
&= \mathbb{P}_1\left\{T \geq -\frac{\log(x)}{\lambda_0}\right\}, \\
&= \exp\left\{\frac{\lambda_1}{\lambda_0}\log(x)\right\},
\end{aligned}
$$

Now consider the distribution of this p-value under the alternative. We have:

$$\begin{aligned}
\mathbb{P}_1(P \leq x) &= \mathbb{P}_1(e^{-\lambda_0 T} \leq x), \\
&= \mathbb{P}_1\{-\lambda_0 T \leq \log(x)\}, \\
&= \mathbb{P}_1\left\{T \geq -\frac{\log(x)}{\lambda_0}\right\}, \\
&= \exp\left\{\frac{\lambda_1}{\lambda_0}\log(x)\right\}, \\
&= x^{\lambda_1/\lambda_0},
\end{aligned}$$

Now consider the distribution of this p-value under the alternative. We have:

$$
\begin{aligned}
\mathbb{P}_1(P \leq x) &= \mathbb{P}_1(e^{-\lambda_0 T} \leq x), \\
&= \mathbb{P}_1\{-\lambda_0 T \leq \log(x)\}, \\
&= \mathbb{P}_1\left\{T \geq -\frac{\log(x)}{\lambda_0}\right\}, \\
&= \exp\left\{\frac{\lambda_1}{\lambda_0}\log(x)\right\}, \\
&= x^{\lambda_1/\lambda_0},
\end{aligned}
$$

so that the distribution of P under H_1 is

$$
f(x) = \frac{\lambda_1}{\lambda_0} x^{\frac{\lambda_1}{\lambda_0}-1},
$$

which is the density of a $\mathrm{Beta}(\lambda_1/\lambda_0, 1)$ distribution.

这里这个p值是满足beta分布

Fisher's method is by far the most common method for combining p-values but consider this thought-provoking result.

Remember we are in the hypothesis testing framework:

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0,1],$$

versus,

$$H_1 : P_i \overset{ind}{\sim} F_i, \text{density } f_i.$$

Admissibility criterion (Birnbaum, 1954)

If H_0 is rejected for any given set of p_i, then it will also be rejected for all sets of p_i^* such that $p_i^* \leq p_i$ for each i.

Fisher's method is by far the most common method for combining p-values but consider this thought-provoking result.

Remember we are in the hypothesis testing framework:

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0, 1],$$

versus,

$$H_1 : P_i \overset{ind}{\sim} F_i, \text{density } f_i.$$

Admissibility criterion (Birnbaum, 1954)

If H_0 is rejected for any given set of p_i, then it will also be rejected for all sets of p_i^* such that $p_i^* \leq p_i$ for each i.

Theorem (Birnbaum, 1954)

For each method of combination satisfying the admissibility criterion, we can find some alternative H_1 represented by non-increasing densities f_i against which that method of combination gives a best test of H_0.

Fisher's method is by far the most common method for combining p-values but consider this thought-provoking result.

Remember we are in the hypothesis testing framework:

$$H_0 : P_i \overset{i.i.d.}{\sim} \text{uniform}[0, 1],$$

$$\text{versus},$$

$$H_1 : P_i \overset{ind}{\sim} F_i, \text{density } f_i.$$

Admissibility criterion (Birnbaum, 1954)

If H_0 is rejected for any given set of p_i, then it will also be rejected for all sets of p_i^* such that $p_i^* \leq p_i$ for each i.

不是global 的了?

Theorem (Birnbaum, 1954)

For each method of combination satisfying the admissibility criterion, we can find some alternative H_1 represented by non-increasing densities f_i against which that method of combination gives a best test of H_0.

$\Rightarrow$ In this sense, there isn't a well-defined "best" combiner.

Let $P_{(1)} \leq \ldots \leq P_{(n)}$ denote the order statistics of the p-values.

We will temporarily reverse the convention and assume that we are rejecting for small statistics T.

According to Birnbaum the following combiners are admissible/inadmissible:

$$T = \min(P_i) = P_{(1)},$$
$$T = \max(P_i) = P_{(n)},$$
$$T = \min\left\{\frac{nP_{(i)}}{i}\right\}.$$

哪里看出来是best test了?

Let $P_{(1)} \leq \ldots \leq P_{(n)}$ denote the order statistics of the p-values.

We will temporarily reverse the convention and assume that we are rejecting for small statistics T.

According to Birnbaum the following combiners are admissible/inadmissible:

$$
\begin{aligned}
T &= \min(P_i) = P_{(1)}, && \checkmark \\
T &= \max(P_i) = P_{(n)}, \\
T &= \min\left\{\frac{nP_{(i)}}{i}\right\}.
\end{aligned}
$$

Let $P_{(1)} \leq \ldots \leq P_{(n)}$ denote the order statistics of the p-values.

We will temporarily reverse the convention and assume that we are rejecting for small statistics T.

According to Birnbaum the following combiners are admissible/inadmissible:

$$
\begin{aligned}
T &= \min(P_i) = P_{(1)}, && \checkmark \\
T &= \max(P_i) = P_{(n)}, && \checkmark \\
T &= \min\left\{\frac{nP_{(i)}}{i}\right\}.
\end{aligned}
$$

Let $P_{(1)} \leq \ldots \leq P_{(n)}$ denote the order statistics of the p-values.

We will temporarily reverse the convention and assume that we are rejecting for small statistics T.

According to Birnbaum the following combiners are admissible:

$$T = \min(P_i) = P_{(1)}, \qquad \checkmark$$
$$T = \max(P_i) = P_{(n)}, \qquad \checkmark$$
$$T = \min\left\{\frac{nP_{(i)}}{i}\right\}. \qquad \checkmark$$

And so his result means that each is best for some alternative.

Distribution of the minimum

Under the null hypothesis

$$\mathbb{P}\{\min(P_i) \geq x\} = \mathbb{P}(P_1, \ldots, P_n \geq x)$$

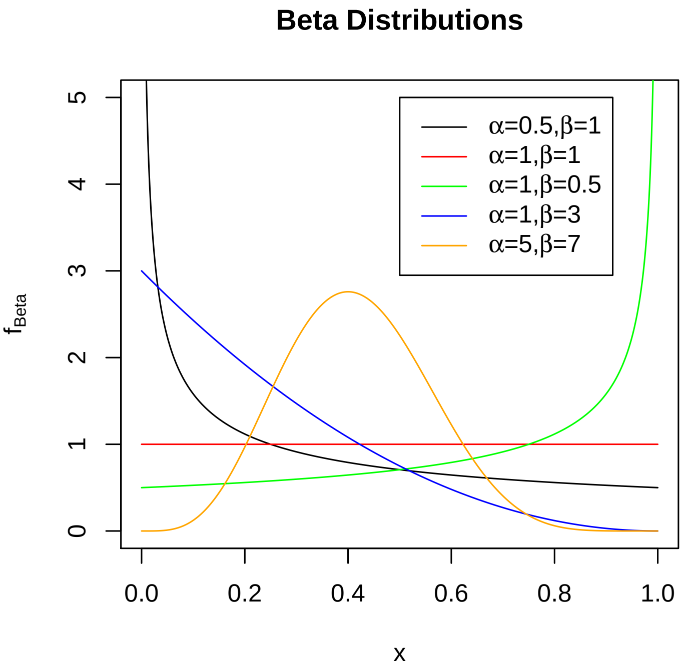
Distribution of the minimum

Under the null hypothesis

$$\begin{aligned}\mathbb{P}\{\min(P_i) \geq x\} &= \mathbb{P}(P_1, \ldots, P_n \geq x) \\ &= (1-x)^n\end{aligned}$$

Thus, $\min(P_i)$ has a density proportional to $(1-x)^{n-1}$ and is therefore $\text{Beta}(1, n)$.

By a similar argument, $\max(P_i)$ is $\text{Beta}(n, 1)$.



Beta Distributions
α=0.5,β=1
α=1,β=1
α=1,β=0.5
α=1,β=3
α=5,β=7
f_{Beta}
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
x

Simes' Method

The combiner

$$T = \min\left\{\frac{nP_{(i)}}{i}\right\}$$

is called Simes' method (Simes, 1986) and has a rather stunning distribution under H_0:

$$\min\left\{\frac{nP_{(i)}}{i}\right\} \sim \text{uniform}[0, 1].$$

Simes gives a simple proof (apparently suggested by the referee), but we will do something more elaborate (and more educational).

Lemma

Let U be a uniform random variable on $[0,1]$. Then for $x \in [0,1]$, the random variable $\tilde{U} = [U \mid U \leq x] \sim \textit{uniform}[0,x]$.

Proof.

Remember that a uniform random variable on a set has a constant density, so that the uniform on $[0,x]$ has density ␣

$\square$

Lemma

Let U be a uniform random variable on $[0,1]$. Then for $x \in [0,1]$, the random variable $\tilde{U} = [U \mid U \leq x] \sim \textit{uniform}[0,x]$.

Proof.

Remember that a uniform random variable on a set has a constant density, so that the uniform on $[0,x]$ has density $1/x$.

$\square$

Lemma

Let U be a uniform random variable on $[0,1]$. Then for $x \in [0,1]$, the random variable $\tilde{U} = [U \mid U \leq x] \sim$ uniform$[0,x]$.

Proof.

Remember that a uniform random variable on a set has a constant density, so that the uniform on $[0,x]$ has density $1/x$.

For $a \in [0,x]$,

$$\begin{aligned}
\mathbb{P}(\tilde{U} \leq a) &= \mathbb{P}(U \leq a \mid U \leq x), \\
&= \frac{\mathbb{P}(U \leq a)}{\mathbb{P}(U \leq x)}, \\
&= a/x.
\end{aligned}$$

so that $\tilde{U}$ has a density $1/x$. □

Lemma

Let U be a uniform random variable on $[0,1]$. Then for $x \in [0,1]$, the random variable $\tilde{U} = [U \mid U \leq x] \sim$ uniform$[0,x]$.

Proof.

Remember that a uniform random variable on a set has a constant density, so that the uniform on $[0,x]$ has density $1/x$.

For $a \in [0,x]$,

$$\begin{aligned} \mathbb{P}(\tilde{U} \leq a) &= \mathbb{P}(U \leq a \mid U \leq x), \\ &= \frac{\mathbb{P}(U \leq a)}{\mathbb{P}(U \leq x)}, \\ &= a/x. \end{aligned}$$

so that $\tilde{U}$ has a density $1/x$. □

$\Rightarrow$ Remember more generally that if we condition on a uniform(A) being in a set $B \subseteq A$, the conditioned variable is uniform(B).

A semi-obvious lemma

Let $X_{(1:n)} \leq \cdots \leq X_{(n:n)}$ denote order statistics corresponding to $i.i.d.$ replicates $X_1, \ldots, X_n$ of a random variable X.

Lemma

Conditional on $X_{(k+1:n)} = x$, the order statistics $X_{(1:n)}, \ldots, X_{(k:n)}$ are distributed as the order statistics $\tilde{X}_{(1:k)} \leq \cdots \leq \tilde{X}_{(k:k)}$ of i.i.d. replicates $\tilde{X}_1, \ldots, \tilde{X}_k$ of the random variable $\tilde{X} = [X \mid X \leq x]$, and are independent of $X_{(k+2:n)}, \ldots, X_{(n:n)}$.

Example

Let $U_{(1:n)} \leq \cdots \leq U_{(n:n)}$ denote the order statistics from a uniform$[0,1]$ distribution.

Then conditional on $U_{(k+1:n)}$, the random variables $U_{(1:n)}, \ldots, U_{(k:n)}$ are distributed as the order statistics $\tilde{U}_{(1:k)} \leq \cdots \leq \tilde{U}_{(k:k)}$ of i.i.d. replicates $\tilde{U}_1, \ldots, \tilde{U}_k$ of

$$\tilde{U} = [U \mid U \leq U_{(k+1:n)}] \sim \text{uniform}[0, U_{(k+1:n)}],$$

and are independent of $U_{(k+2:n)}, \ldots, U_{(n:n)}$.

Example

Let $U_{(1:n)} \leq \cdots \leq U_{(n:n)}$ denote the order statistics from a uniform$[0, 1]$ distribution.

Then conditional on $U_{(k+1:n)}$, the random variables $U_{(1:n)}, \ldots, U_{(k:n)}$ are distributed as the order statistics $\tilde{U}_{(1:k)} \leq \cdots \leq \tilde{U}_{(k:k)}$ of i.i.d. replicates $\tilde{U}_1, \ldots, \tilde{U}_k$ of

$$\tilde{U} = [U \mid U \leq U_{(k+1:n)}] \sim \text{uniform}[0, U_{(k+1:n)}],$$

and are independent of $U_{(k+2:n)}, \ldots, U_{(n:n)}$.

Therefore, the random variables $U_{(1:n)}/U_{(k+1:n)}, \ldots, U_{(k:n)}/U_{(k+1:n)}$ are distributed as the k order statistics from a uniform$[0, 1]$ distribution, and independent of $U_{(k+1:n)}, \ldots, U_{(n:n)}$.

A toy application

It therefore follows that:

$$-2\log\{U_{(1:n)}\} + 2\log\{U_{(2:n)}\} \sim \text{␣}$$

It therefore follows that:

$$-2\log\{U_{(1:n)}\} + 2\log\{U_{(2:n)}\} \sim \chi^2_2.$$

Strange ways of simulating a uniform random variable

Remember $U_{(1:n)} \leq \cdots \leq U_{(n:n)}$ are the order statistics from a uniform$[0, 1]$ distribution.

Let:

$$Y = U_{(Z:n)}, \quad \text{where } Z \sim \text{categorical}\left(\frac{1}{n}, \cdots, \frac{1}{n}\right) = \text{multinomial}\left(\frac{1}{n}, \cdots, \frac{1}{n}; 1\right).$$

then:

$$Y \sim \text{␣}$$

Strange ways of simulating a uniform random variable

Remember $U_{(1:n)} \leq \cdots \leq U_{(n:n)}$ are the order statistics from a uniform$[0,1]$ distribution.

Let:

$$Y = U_{(Z:n)}, \quad \text{where } Z \sim \text{categorical}\left(\frac{1}{n}, \cdots, \frac{1}{n}\right) = \text{multinomial}\left(\frac{1}{n}, \cdots, \frac{1}{n}; 1\right).$$

then:

$$Y \sim \text{uniform}[0,1].$$

Strange ways of simulating a uniform random variable

Equivalently, we'll write

$$Y = U_{(n:n)}X,$$

where

$$X = \frac{U_{(Z:n)}}{U_{(n:n)}}, \quad \text{and } Z \sim \text{categorical}\left(\frac{1}{n}, \cdots, \frac{1}{n}\right),$$

so that as before $Y \sim \text{uniform}[0, 1]$.

Then $X = 1$ with probability $1/n$, $X < 1$ with probability $(n-1)/n$, and

$$[X \mid X < 1] = \frac{U_{(Z:n)}}{U_{(n:n)}}, \quad \text{where } Z \sim \text{categorical}\left(\frac{1}{n-1}, \cdots, \frac{1}{n-1}\right) \text{ on } \{1, \ldots, n-1\},$$

so that

$$[X \mid X < 1] \sim \text{␣}$$

Equivalently, we'll write

$$Y = U_{(n:n)} X,$$

where

$$X = \frac{U_{(Z:n)}}{U_{(n:n)}}, \quad \text{and } Z \sim \text{categorical}\left(\frac{1}{n}, \cdots, \frac{1}{n}\right),$$

so that as before $Y \sim \text{uniform}[0,1]$.

Then $X = 1$ with probability $1/n$, $X < 1$ with probability $(n-1)/n$, and

$$[X \mid X < 1] = \frac{U_{(Z:n)}}{U_{(n:n)}}, \quad \text{where } Z \sim \text{categorical}\left(\frac{1}{n-1}, \cdots, \frac{1}{n-1}\right) \text{ on } \{1, \ldots, n-1\},$$

so that

$$[X \mid X < 1] \sim \text{uniform}[0,1].$$

Strange ways of simulating a uniform random variable

In other words, the following (pointless) random mechanism generates a uniform random variable on $[0, 1]$:

1. Simulate $U_{(n:n)} \sim \text{Beta}(n, 1)$.

In other words, the following (pointless) random mechanism generates a uniform random variable on $[0,1]$:

1. Simulate $U_{(n:n)} \sim \text{Beta}(n,1)$.
2. Accept and return $U_{(n:n)}$ with probability $1/n$.

In other words, the following (pointless) random mechanism generates a uniform random variable on $[0, 1]$:

1. Simulate $U_{(n:n)} \sim \text{Beta}(n, 1)$.
2. Accept and return $U_{(n:n)}$ with probability $1/n$.
3. Otherwise, generate $X^* = [X \mid X < 1] \sim \text{uniform}[0, 1]$ and return $U_{(n:n)}X^*$ (which simply amounts to simulating a uniform$[0, U_{(n:n)}]$).

Proof of Simes' result

For $n = 1$ we obviously have

$$\min\left\{\frac{nU_{(i:n)}}{i}\right\} \sim \text{uniform}[0, 1].$$

Now suppose the result holds for $n - 1$.

Proof of Simes' result

We have

$$\min\left\{\frac{nU_{(i:n)}}{i}\right\} = \min\left\{\frac{nU_{(1:n)}}{1}, \ldots, \frac{nU_{(n:n)}}{n}\right\},$$

Proof of Simes' result

We have

$$
\begin{aligned}
\min\left\{\frac{nU_{(i:n)}}{i}\right\} &= \min\left\{\frac{nU_{(1:n)}}{1},\dots,\frac{nU_{(n:n)}}{n}\right\}, \\
&= U_{(n:n)}\min\left\{\frac{nU_{(1:n)}/U_{(n:n)}}{1},\dots,1\right\},
\end{aligned}
$$

Proof of Simes' result

We have

$$\begin{aligned}
\min\left\{\frac{nU_{(i:n)}}{i}\right\} &= \min\left\{\frac{nU_{(1:n)}}{1},\dots,\frac{nU_{(n:n)}}{n}\right\}, \\
&= U_{(n:n)}\min\left\{\frac{nU_{(1:n)}/U_{(n:n)}}{1},\dots,1\right\}, \\
&= U_{(n:n)}\min\left[\min\left\{\frac{nU_{(1:n)}/U_{(n:n)}}{1},\dots,\frac{nU_{(n-1:n)}/U_{(n:n)}}{n-1}\right\},1\right],
\end{aligned}$$

We have

$$
\begin{aligned}
\min\left\{\frac{nU_{(i:n)}}{i}\right\} &= \min\left\{\frac{nU_{(1:n)}}{1},\ldots,\frac{nU_{(n:n)}}{n}\right\}, \\
&= U_{(n:n)}\min\left\{\frac{nU_{(1:n)}/U_{(n:n)}}{1},\ldots,1\right\}, \\
&= U_{(n:n)}\min\left[\min\left\{\frac{nU_{(1:n)}/U_{(n:n)}}{1},\ldots,\frac{nU_{(n-1:n)}/U_{(n:n)}}{n-1}\right\},1\right], \\
&= U_{(n:n)}\min\left[\frac{n}{n-1}\min\left\{\frac{(n-1)U_{(1:n)}/U_{(n:n)}}{1},\ldots,\frac{(n-1)U_{(n-1:n)}/U_{(n:n)}}{n-1}\right\},1\right],
\end{aligned}
$$

Proof of Simes' result

We have

$$\begin{aligned}
\min\left\{\frac{nU_{(i:n)}}{i}\right\} &= \min\left\{\frac{nU_{(1:n)}}{1},\ldots,\frac{nU_{(n:n)}}{n}\right\},\\
&= U_{(n:n)}\min\left\{\frac{nU_{(1:n)}/U_{(n:n)}}{1},\ldots,1\right\},\\
&= U_{(n:n)}\min\left[\min\left\{\frac{nU_{(1:n)}/U_{(n:n)}}{1},\ldots,\frac{nU_{(n-1:n)}/U_{(n:n)}}{n-1}\right\},1\right],\\
&= U_{(n:n)}\min\left[\frac{n}{n-1}\min\left\{\frac{(n-1)U_{(1:n)}/U_{(n:n)}}{1},\ldots,\frac{(n-1)U_{(n-1:n)}/U_{(n:n)}}{n-1}\right\},1\right],\\
&\sim U_{(n:n)}\min\left\{\frac{n}{n-1}\tilde{U},1\right\},
\end{aligned}$$

where $\tilde{U} \sim \text{uniform}[0,1]$ independently (by the recursion assumption).

Proof of Simes' result

Let $X = \min\left\{\frac{n}{n-1}\tilde{U}, 1\right\}$. Then $X = 1$ with probability ␣

Proof of Simes' result

Let $X = \min\left\{\frac{n}{n-1}\tilde{U}, 1\right\}$. Then $X = 1$ with probability $1/n$.

Therefore $X < 1$ with probability $(n-1)/n$ and

$$[X \mid X < 1] \sim \text{uniform}[0,1],$$

since

$$[X \mid X < 1] = \left[\frac{n}{n-1}\tilde{U} \;\middle|\; \frac{n}{n-1}\tilde{U} < 1\right],$$

i.e. is a uniform random variable on $[0, n/(n-1)]$ conditioned on being smaller than 1.

Proof of Simes' result

Let $X = \min\left\{\frac{n}{n-1}\tilde{U}, 1\right\}$. Then $X = 1$ with probability $1/n$.

Therefore $X < 1$ with probability $(n-1)/n$ and

$$[X \mid X < 1] \sim \text{uniform}[0, 1],$$

since

$$[X \mid X < 1] = \left[\frac{n}{n-1}\tilde{U} \;\middle|\; \frac{n}{n-1}\tilde{U} < 1\right],$$

i.e. is a uniform random variable on $[0, n/(n-1)]$ conditioned on being smaller than 1.

Putting it all together, we have:

$$\begin{aligned}\min\left\{\frac{nU_{(i:n)}}{i}\right\} &\sim U_{(n:n)} \min\left\{\frac{n}{n-1}\tilde{U}, 1\right\},\\ &\sim U_{(n:n)}X,\end{aligned}$$

Let $X = \min\left\{\frac{n}{n-1}\tilde{U}, 1\right\}$. Then $X = 1$ with probability $1/n$.

Therefore $X < 1$ with probability $(n-1)/n$ and

$$[X \mid X < 1] \sim \text{uniform}[0,1],$$

since

$$[X \mid X < 1] = \left[\frac{n}{n-1}\tilde{U} \;\middle|\; \frac{n}{n-1}\tilde{U} < 1\right],$$

i.e. is a uniform random variable on $[0, n/(n-1)]$ conditioned on being smaller than 1.

Putting it all together, we have:

$$\begin{aligned}
\min\left\{\frac{nU_{(i:n)}}{i}\right\} &\sim U_{(n:n)} \min\left\{\frac{n}{n-1}\tilde{U}, 1\right\}, \\
&\sim U_{(n:n)} X, \\
&\sim \text{uniform}[0,1].
\end{aligned}$$

Simes' proof

752 R. J. SIMES

THEOREM. *Let $P_{(1)}, \ldots, P_{(n)}$ be the order statistics of n independent uniform* (0, 1) *random variables and let* $A_n(\alpha) = \mathrm{pr}\{P_{(j)} > j\alpha/n; j = 1, \ldots, n\}$ $(0 \leqslant \alpha \leqslant 1)$. *Then* $A_n(\alpha) = 1 - \alpha$.

Proof. The result is clearly true for $n = 1$. For $n > 1$, $\{P_{(1)}/P_{(n)}, \ldots, P_{(n-1)}/P_{(n)}\}$ are the order statistics of $n-1$ independent uniform random variables on (0, 1), independent of $P_{(n)}$, and $P_{(n)}$ has distribution function p^n $(0 < p < 1)$. Hence

$$A_n(\alpha) = \int_\alpha^1 A_{n-1}\left\{\frac{\alpha(n-1)}{pn}\right\} np^{n-1}\, dp.$$

If $A_{n-1}(\alpha) = 1 - \alpha$ then $A_n(\alpha) = 1 - \alpha$ follows. Hence the result is proved by induction. $\square$

A comparison

- Remember that with ten p-values respectively equal to 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.01, 0.01, 0.01, we obtain $-2\sum \log(p_i) \approx 37$, $1 - F_{\chi^2}(37, 20) \approx 0.01$.
- For the same p-values we obtain $\min(p_i) = 0.01$, $F_{\mathrm{Beta}(1,n)}(0.01) \approx 0.09561792$.
- and $\max(p_i) = 0.5$, $F_{\mathrm{Beta}(n,1)}(0.5) \approx 0.001$
- and finally (the test-statistic is the p-value!):

$$\min\left\{\frac{np_{(i)}}{i}\right\} \approx 0.03.$$

Real data example

- Authentication event host logs obtained from the Los Alamos National Laboratory (LANL) computer network (Turcotte et al., 2017), collected from all LANL computers running the Microsoft Windows operating system.
- The data consist of over 10,000 user accounts authenticating on over 15,000 different computers over a period of 90 days.
- For each event record in these data, a username and event ID is provided; the latter indicates what type of authentication event occurred, such as a network log on, a workstation lock, or an interactive log on.

Authentication events

Table 3. Event IDs used from the LANL authentication data.

Event ID	Description
4624	A user successfully logged onto a computer
4625	A user failed to log onto a computer
4634	A user logged off a computer
4800	The workstation was locked
4801	The workstation was unlocked
4802	The screensaver was invoked
4803	The screensaver was dismissed

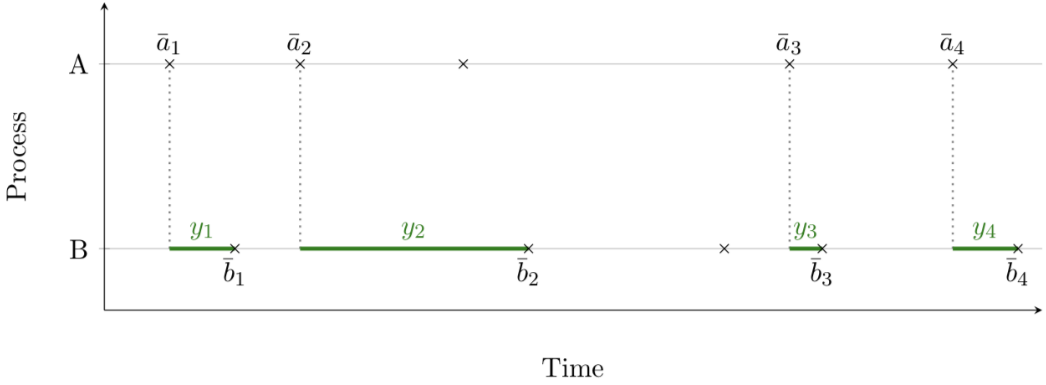
Table 4. Log on types used for event IDs $\{4624, 4625, 4634\}$.

Log on type	Description
2	Interactive
7	Unlock
10	Remote Interactive

- Example from "Detecting weak dependence in computer network traffic patterns using higher criticism" (Price-Williams *et al.*, 2018) (to appear in the *Journal of the Royal Statistical Society*, Series C).
- The paper is generally concerned with connecting two sequences of events on the basis of coincidental timings.
- In IMCS we talked about applications such as detecting whether an email recipient clicked on an embedded URL using corporate email logs (which have information about attachment) and web logs (indicating whether user went to a website), but of course this is a very general problem.

Illustration of methodology

Without going into details, we can convert inter-event times into sequences of p-values quickly and (more or less) effectively.



Example user

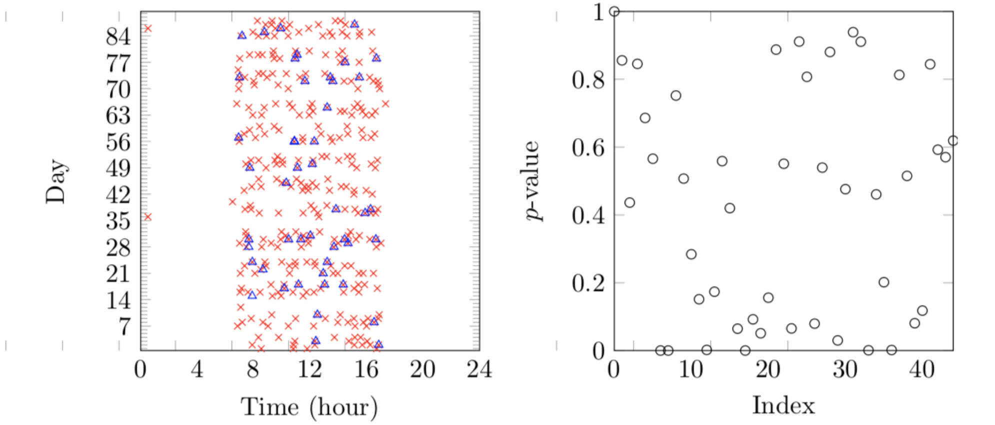


Fig. 6. An example of weak triggering behaviour in authentication data from User 233172 in the LANL network. Left: Times of failed log on events (△) and screen saver disable events (×). Right: Corresponding p-values for the waiting times between screen saver dismissal and failed log on.

We will investigate whether we can find temporal dependence between screensaver dismissal and log on failure events. Reasons:

1. Intuitively, there should be a dependence.
2. However, it is clear that many p-values should be irrelevant (most of the time, the password is entered correctly).

Another example

Pattern	P-value	Time lag	Subject
$11 \to j \rightsquigarrow j \to 11$	9.6×10^{-12}	0:00:00	**DWR - Gas Daily**
		0:00:58	RE: DWR - Gas Daily
		0:58:31	RE: DWR - Gas Daily
		1:06:49	RE: DWR - Gas Daily
		1:27:12	RE: DWR - Gas Daily
$2 \to j \rightsquigarrow j \to 2$	4.7×10^{-9}	0:00:00	**RE: CPUC Questions on DA**
		0:00:19	RE: CPUC Questions on DA
$4 \to j \rightsquigarrow j \to 4$	4.9×10^{-8}	0:00:00	**RE: Transwestern Hearing**
		0:14:00	RE: Transwestern Hearing
		5:01:00	RE: Transwestern Hearing
$12 \to j \rightsquigarrow j \to 12$	1.5×10^{-6}	0:00:00	**RE: CA Unbundling**
		0:04:58	RE: CA Unbundling
$9 \to j \rightsquigarrow j \to 3$	4.6×10^{-5}	0:00:00	**California Update–Legislative Push Underway**
		0:51:00	Re: California Update–Legislative Push Underway
		1:03:00	Re: California Update–Legislative Push Underway
			⋮
		10:51:00	Re: California Update–Legislative Push Underway
		11:03:00	Re: California Update–Legislative Push Underway
			⋮
$1 \to j \rightsquigarrow j \to 1$	4.7×10^{-5}	0:00:00	**Re: Comments to Govś Proposals**
		0:02:00	Re: Comments to Govś Proposals
		5:39:00	RE: Additional Materials
		21:37:00	Update from EES Call this Morning

Another example (cont')

Pattern	P-value	Time lag	Subject
$3 \to j \rightsquigarrow j \to 3$	9.3×10^{-5}	0:00:00	**Re: Pescetti**
		0:03:00	RE: Pescetti
$9 \to j \rightsquigarrow j \to 9$	3.7×10^{-4}	0:00:00	**California Update–Legislative Push Underway**
		0:51:00	Re: California Update–Legislative Push Underway
		10:51:00	Re: California Update–Legislative Push Underway
$2 \to j \rightsquigarrow j \to 6$	8.6×10^{-4}	0:00:00	**HERE IS MY DRAFT**
		0:09:00	Re: FW: SoCalGas Capacity Forum
$6 \to j \rightsquigarrow j \to 6$	1.7×10^{-3}	0:00:00	**Re: FW: SoCalGas Capacity Forum**
		2:22:00	Re: FW: SoCalGas Capacity Forum
$5 \to j \rightsquigarrow j \to 5$	2.1×10^{-3}	0:00:00	**Re: Response to ORA/TURN petition**
		0:03:00	Re: Response to ORA/TURN petition
$8 \to j \rightsquigarrow j \to 5$	2.2×10^{-3}	0:00:00	**RE: Call to Discuss Possible Options to Mitigate Ef...**
		1:18:29	Re:
$4 \to j \rightsquigarrow j \to 10$	2.4×10^{-3}	0:00:00	**RE: Transwestern Hearing**
		1:31:13	Attorneys
$8 \to j \rightsquigarrow j \to 8$	3.8×10^{-3}	0:00:00	**RE: Call to Discuss Possible Options to Mitigate Ef...**
		2:11:29	RE: Call to Discuss Possible Options to Mitigate Ef...
$5 \to j \rightsquigarrow j \to 4$	7.1×10^{-3}	0:00:00	**FW: EPSA report**
		41:33:00	RE: Transwestern Hearing
		46:20:00	RE: Transwestern Hearing
$2 \to j \rightsquigarrow j \to 11$	7.5×10^{-3}	0:00:00	**Willie Brown INFO**
		0:03:48	RE: Socal Storage Projects
$7 \to j \rightsquigarrow j \to 1$	8.4×10^{-3}	0:00:00	**Governor DavisṔress conference Highlights – wil...**
		4:23:00	
		5:06:00	Email for Transmittal from Ken Lay to Senator Brult...
$6 \to j \rightsquigarrow j \to 5$	9.1×10^{-3}	0:00:00	**Re: FW: SoCalGas Capacity Forum**
		2:22:00	Re: FW: SoCalGas Capacity Forum

A harder problem

- A result from "Higher criticism for detecting sparse heterogeneous mixtures" (Donoho and Jin), *Annals of Statistics, 2004*.
- The hypothesis testing framework considered is:

$$H_0 : X_1, \dots, X_n \overset{i.i.d}{\sim} \text{normal}(0, 1),$$

versus,

$$H_1 : X_1, \dots, X_n \overset{i.i.d}{\sim} \epsilon_n \text{normal}(\mu_n, 1) + (1 - \epsilon_n)\text{normal}(0, 1).$$

- They set $\epsilon_n = n^{-\beta}$ with $\beta \in (1/2, 1)$ (so that the proportion of anomalies goes to zero but the number of anomalies still keeps growing), and choose μ_n to make the problem *very* hard.
- In particular, there is a boundary $b_n = \sqrt{2\rho(\beta)\log(n)}$ (where $\rho(\beta) \in (0, 1)$ is explicit in the paper) such that if $\mu_n > b_n$ the likelihood ratio test is consistent, i.e. will asymptotically reject with probability one ($\beta = 1$), whereas if $\mu_n < b_n$, the likelihood ratio test is asymptotically unable to distinguish the hypotheses ($\beta = \alpha$).

Their result

- Fisher's method is asymptotically unable to distinguish the hypotheses ($\beta = \alpha$).

Their result

- Fisher's method is asymptotically unable to distinguish the hypotheses ($\beta = \alpha$).
- Asymptotically, the statistic:

$$\sqrt{n}\max\left[\{i/n - p_{(i)}\}/\sqrt{p_{(i)}(1 - p_{(i)})}\right],$$

is consistent ($\beta = 1$) whenever the likelihood ratio test is consistent.

Their result

- Fisher's method is asymptotically unable to distinguish the hypotheses ($\beta = \alpha$).
- Asymptotically, the statistic:

$$\sqrt{n} \max \left[\{ i/n - p_{(i)} \} / \sqrt{p_{(i)}(1 - p_{(i)})} \right],$$

  is consistent ($\beta = 1$) whenever the likelihood ratio test is consistent.
- This is both interesting and useful, since likelihood ratio test is not computable in practice, and is not necessarily consistent if the wrong ϵ_n and μ_n are used.