A Bayesian approach to power law analysis

Luiz Max F. de Carvalho
Program for Scientific Computing (PROCC), Oswaldo Cruz Foundation.

January 8, 2019

1 Ideas

Taking the method of Clauset et al. (2009) as a basis:

- Replace bootstrap p-value with posterior predictive p-value (Gabry et al., 2017) for assessing adequacy of power law: need an experiment showing comparable or better performance.
- Replace likelihood ratio tests with Bayes factors: need experiment showing BFs are comparable or better than LR tests

  Marginal likelihoods (MaL) are "easy" for power law and stretched exponential, need to figure out if log-normal can be done analytically. Marginal likelihood for PL can be made more general by assuming a Gamma(a_β, b_β) prior on β – need to talk about the induced density on $\alpha = \beta + 1$. To recover the Jeffreys prior just set $a_\beta = b_\beta = 0$[1]. Need to check the results of MaL for stretched exponential: some discrepancy with bridge sampling results. Should not there be any. I think this can be tracked down to instability with `lintegrate`.

  Sketch of experiment to test BFs: simulate 100 (1000) data sets from a power law as in the simulated example in Clauset et al, 100 data sets from a log-normal with same mean and variance (aka moment-matching) and 100 data sets from a moment matching stretched exponential. Then perform LR tests and BFs and count how many times each procedure picked the right model. Notice this falls in the category of "frequentist properties of Bayesian estimators". But if we are to have any hope of convincing physicists to switch approaches we will need to show that a Bayesian approach gets the answer right every time the frequentist one does and also in some situations where frequentism fails.

2 Some calculations

For the simplest power law model, the likelihood is

$$f(x|\alpha, x_{\text{min}}) = \frac{\alpha - 1}{x_{\text{min}}} \left(\frac{x}{x_{\text{min}}}\right)^{-\alpha}.$$

When a vector of observations $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$ is available, the likelihood is

$$f(\boldsymbol{x}|\alpha, x_{\text{min}}) = \prod_{i=1}^{N} f(x_i|\alpha, x_{\text{min}}) = \left(\frac{\alpha - 1}{x_{\text{min}}}\right)^{N} \left(\frac{\prod_{i=1}^{N} x_i}{x_{\text{min}}^{N}}\right)^{-\alpha},$$

under the assumption of independent and identically-distributed data.

The Jeffreys prior (Jeffreys, 1946) for α is

$$\pi_J(\alpha) \propto \frac{1}{\alpha - 1},$$

and leads to a proper posterior distribution. Moreover, when x_{min} is assumed known, the marginal

[1] Mathematically different,numerically the same. See https://stats.stackexchange.com/a/111227/97431

likelihood with respect to α can be computed analytically:

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}|x_{\min}) &= \int_1^\infty \frac{1}{(\alpha-1)}\left(\frac{\alpha-1}{x_{\min}}\right)^N \left(\frac{\prod_{i=1}^N x_i}{x_{\min}^N}\right)^{-\alpha} d\alpha, \\ &= \frac{1}{\prod_{i=1}^N x_i}\int_0^\infty \left(\frac{x_{\min}^N}{\prod_{i=1}^N x_i}\right)^\beta \beta^{N-1} d\beta, \\ &= \frac{1}{\prod_{i=1}^N x_i}\frac{\Gamma(N)}{\left(-\log\left(\frac{x_{\min}^N}{\prod_{i=1}^N x_i}\right)\right)^N}.\end{aligned}$$

Could also employ a Gamma(a_β, b_β) on $\beta = \alpha - 1$ and the marginal likelihood would be:

$$\mathcal{L}(\boldsymbol{x}|x_{\min}, a_\beta, b_\beta) = \frac{1}{\prod_{i=1}^N x_i}\frac{\Gamma(N+a_\beta)}{\left(-\log\left(\frac{x_{\min}^N}{\prod_{i=1}^N x_i}\right)+b_\beta\right)^{N+a_\beta}}. \tag{1}$$

The problem with this approach is that the induced prior on the actual parameter of interest, α, is a bit weird.

Accommodating truncated priors

Gillespie et al. (2017) propose an uniform prior on α with bounds $l_\alpha = 3/2$ and $u_\alpha = 3$. In this situation we have

$$\mathcal{L}(\boldsymbol{x}|x_{\min}, l_\alpha, u_\alpha) = \int_{l_\alpha}^{u_\alpha} \frac{1}{(u_\alpha - l_\alpha)}\left(\frac{\alpha-1}{x_{\min}}\right)^N \left(\frac{\prod_{i=1}^N x_i}{x_{\min}^N}\right)^{-\alpha} d\alpha, \tag{2}$$

$$= \frac{\Gamma\left(N+1, -(l_\alpha - 1)\log(z)\right) - \Gamma\left(N+1, -(u_\alpha-1)\log(z)\right)}{(u_\alpha - l_\alpha)\prod_{i=1}^N x_i \left(-\log(z)\right)^N \log(z)}, \tag{3}$$

for $1 \leq l_\alpha < u_\alpha$ with $z := \frac{x_{\min}^N}{\prod_{i=1}^N x_i}$, where $\Gamma(s, x)$ is the incomplete Gamma function and $\Gamma(s) = \Gamma(s, 0)$.

3 Log-normal

The likelihood is:

$$f(\boldsymbol{x}|\mu, \sigma, x_{\min}) = \prod_{i=1}^n f(x_i|\mu, \sigma, x_{\min}), \tag{4}$$

$$= \left(\prod_{i=1}^n x_i\right)^{-1}\left(\prod_{i=1}^n x_i\right)^{\frac{\mu}{\sigma^2}}\left(\sqrt{\frac{2}{\pi\sigma^2}}\frac{1}{\operatorname{erfc}\left(\frac{\log x_{\min}-\mu}{\sqrt{2}\sigma}\right)}\right)^n \exp\left(-\frac{n\mu^2}{2\sigma^2}\right)\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(\log x_i)^2\right) \tag{5}$$

Let us first integrate with respect to σ. For that we will make the substitution $\tau := 1/\sigma^2$ and place a Gamma(α_s, β_s) on τ. Also let $\boldsymbol{p} := \prod_{i=1}^n x_i$ and $\boldsymbol{S}_2 := \sum_{i=1}^n (\log x_i)^2$. First, re-arrange the likelihood:

$$f(\boldsymbol{x}|\mu, \sigma, x_{\min}) = \boldsymbol{p}^{-1}\boldsymbol{p}^{\tau\mu}\left(\sqrt{\frac{2\tau}{\pi}}\frac{1}{\operatorname{erfc}\left(\frac{(\log x_{\min}-\mu)\sqrt{\tau}}{\sqrt{2}}\right)}\right)^n \exp\left(-\frac{n\mu^2\tau}{2}\right)\exp\left(-\frac{\tau}{2}\boldsymbol{S}_2\right), \tag{6}$$

$$= \left(\sqrt{\frac{2}{\pi}}\right)^n \boldsymbol{p}^{-1}\left(\boldsymbol{p}^\mu \exp\left(-\frac{n\mu^2 + \boldsymbol{S}_2}{2}\right)\right)^\tau \tau^{n/2}\left(\frac{1}{\operatorname{erfc}\left(\frac{(\log x_{\min}-\mu)\sqrt{\tau}}{\sqrt{2}}\right)}\right)^n \tag{7}$$

Seems like a dead-end. Could try integration by parts with $u = 1/(\operatorname{erfc}(c\sqrt{\tau}))^n$ and $dv = \exp(-A\tau)\tau^{n/2}$. Under the Jeffreys prior $\pi(\tau) \propto \tau$, $dv = \exp(-A\tau)\tau^{(n+2)/2}$. The full plan for a semi-analytical solution here is: integrate one of the parameters out then marginalise over the other one through quadrature (see stretched exponential below).

4 Stretched exponential (Weibull)

The likelihood for a single point is

$$f(x|\lambda, \beta, x_{\min}) = \lambda\beta \exp\left(\lambda x_{\min}^{\beta}\right) x^{\beta-1} \exp\left(-\lambda x^{\beta}\right)$$

The likelihood of a sample $\boldsymbol{x}$ is

$$f(\boldsymbol{x}|\mu, \sigma, x_{\min}) = \prod_{i=1}^{N} f(x_i|\lambda, \beta, x_{\min}), \tag{8}$$

$$= \left(\lambda\beta \exp\left(\lambda x_{\min}^{\beta}\right)\right)^{N} \boldsymbol{p}^{\beta-1} \exp\left(-\lambda \sum_{i=1}^{N} x_i^{\beta}\right). \tag{9}$$

If we then decide to employ Gamma priors for both parameters we get the posterior

$$p(\lambda, \beta|\boldsymbol{x}) \propto \left(\lambda\beta \exp\left(\lambda x_{\min}^{\beta}\right)\right)^{N} \boldsymbol{p}^{\beta-1} \exp\left(-\lambda \sum_{i=1}^{N} x_i^{\beta}\right) \frac{b_1^{a_1}}{\Gamma(a_1)} \beta^{a_1-1} \exp(-b_1\beta) \frac{b_2^{a_2}}{\Gamma(a_2)} \lambda^{a_2-1} \exp(-b_2\lambda) \tag{10}$$

To compute the marginal likelihood we will first integrate the posterior with respect to λ. For that we will collect only terms that depend on λ and get

$$p(\lambda|\boldsymbol{x}, \beta) \propto \lambda^{N} \lambda^{a_2-1} \exp\left(\lambda N x_{\min}^{\beta}\right) \exp\left(-\lambda \sum_{i=1}^{N} x_i^{\beta}\right) \exp(-b_2\lambda),$$

$$\propto \lambda^{N+a_2-1} \exp\left(-\left[\sum_{i=1}^{N} x_i^{\beta} - N x_{\min}^{\beta} + b_2\right]\lambda\right),$$

which is the kernel of a Gamma$(N + a_2, \sum_{i=1}^{N} x_i^{\beta} - N x_{\min}^{\beta} + b_2)$ distribution, leading to

$$p(\beta|\boldsymbol{x}) = \frac{b_1^{a_1}}{\Gamma(a_1)} \frac{b_2^{a_2}}{\Gamma(a_2)} \beta^{N+a_1-1} \boldsymbol{p}^{\beta-1} \exp(-b_1\beta) \frac{\Gamma(N + a_2)}{\left[\sum_{i=1}^{N} x_i^{\beta} - N x_{\min}^{\beta} + b_2\right]^{N+a_2}}, \tag{11}$$

$$= \frac{b_1^{a_1}}{\Gamma(a_1)} \frac{b_2^{a_2}}{\Gamma(a_2)} \frac{\Gamma(N + a_2)}{\boldsymbol{p}} \frac{\beta^{N+a_1-1} \exp\left(-[b_1 - \sum_{i=1}^{N} \log(x_i)]\beta\right)}{\left[\sum_{i=1}^{N} x_i^{\beta} - N x_{\min}^{\beta} + b_2\right]^{N+a_2}}. \tag{12}$$

The integral $\int_0^{\infty} p(\beta|\boldsymbol{x}) d\beta$ is intractable but can be computed by numerically-stable quadrature. I have used the library `lintegrate` (https://github.com/mattpitkin/lintegrate) to do the quadrature; while not rock solid it gets the job done.

References

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2017). Visualization in bayesian workflow. *arXiv preprint arXiv:1709.01449*.

Gillespie, C. S. et al. (2017). Estimating the number of casualties in the american indian war: a bayesian analysis using the power law distribution. *The Annals of Applied Statistics*, 11(4):2357–2374.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461.