MATH 80601A *Bayesian modelling*

Practice midterm examination solutions

Exam booklet

Examiner: Léo Belzile,

Instructions: The time allotted for the examination is 180 minutes. You may answer in either English or French. No written material may be brought into the examination.

There are a total of 35 marks available in the exam paper, the distribution of which can be found in the right margin in brackets.

You must hand back the **exam booklet** at the end of the examination.

Question 1. Short answer questions

1.1 If X is a continuous random variable with density function $f(x)$, is $f(x_0)$ a probability for given x_0? Justify your answer. [1 pt]

> **Solution:** The integral of the density over the domain is unity, but the only constraint otherwise is positivity. Counterexample is $\mathsf{expo}(\lambda)$, with density $\lambda\exp(-\lambda x)$, which reduces to $\lambda > 0$ at zero.

1.2 **Inversion sampling**: for an absolutely continuous distribution function F with quantile function F^{-1} and $U \sim \mathsf{unif}(0,1)$ uniform, prove that $F^{-1}(U)$ has distribution function F. [2 pts]

> **Solution:** The distribution function $F(X) \sim \mathsf{unif}(0,1)$, so applying the monotone transformation F^{-1} on either sides gives $F^{-1}\{F(X)\} = X = F^{-1}(U)$, since the quantile function is the inverse of F.

1.3 Give an example of a distribution $p(Y \mid \theta)$ which may admit as conjugate prior for θ [2 pts]

i. the gamma distribution,

ii. the beta distribution.

> **Solution:** Poisson distribution with gamma prior, binomial distribution with beta prior. For example, if $Y \sim \mathsf{Poisson}(\theta)$ and $\theta \sim \mathsf{gamma}(a,b)$, then $p(y \mid \theta) \propto \theta^y \exp(-\theta)$ and $p(\theta) \propto \theta^{a-1}\exp(-\theta b)$, then $p(\theta \mid y) \propto \theta^{y+a-1}\exp\{-\theta(1+b)\}$ which is $\mathsf{gamma}(y+a, 1+b)$.

Question 2. Modelling unemployment rates

We consider a times series of $n = 142$ monthly harmonised unemployment rates (in percentage) of women in Norway for the period running from April 2000 to January 2012. We treat the sample $\boldsymbol{y}_{1:n} = (y_1, \ldots, y_n)^\top$ as the realization of an autoregressive process of order 1, denoted $\textsc{ar}(1)$, of the form

$$Y_t \mid Y_{t-1} = y_{t-1} \sim \mathsf{Gauss}(\alpha + \phi y_{t-1}, \nu), \qquad t = 1, \ldots, n; \tag{1}$$

where α is a location parameter, ν is the conditional **variance** and ϕ is the autoregressive parameter. Throughout, we perform inference conditional on the observed initial value y_0 for March 2000, treating the latter as a fixed quantity. Recall that an $\textsc{ar}(1)$ process is stationary if $|\phi| < 1$, with the unconditional variance of $\sigma^2/(1-\phi^2)$.

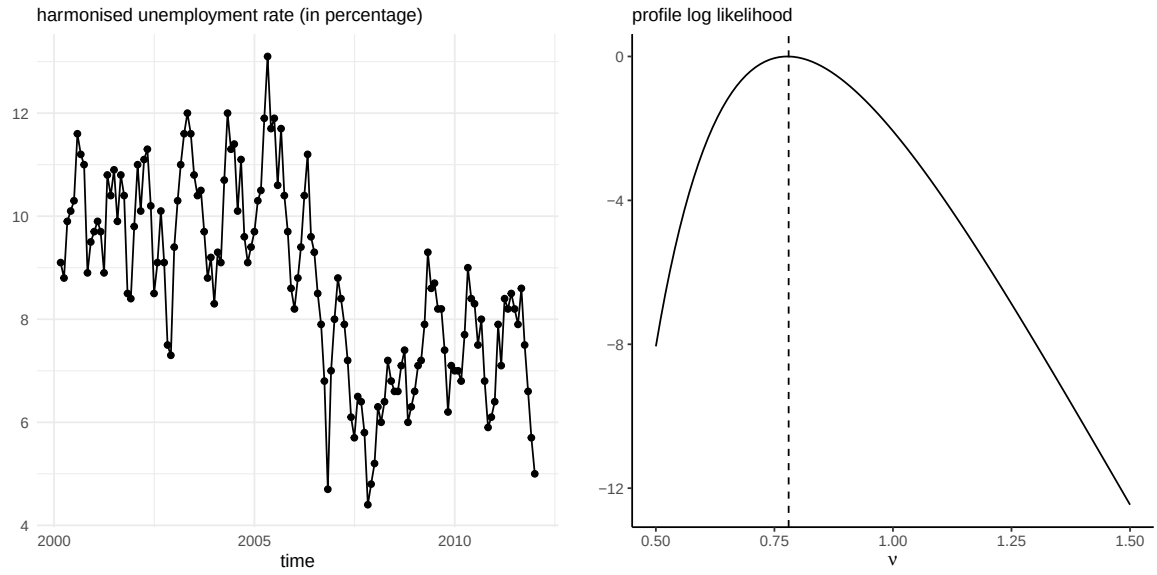


Figure 1: Left: time series of monthly women harmonised unemployment rate over time. Right: Profile log likelihood for the variance parameter ν of eq. (1). The curve has been shifted down so that the profile log likelihood is zero at the maximum likelihood estimate. The vertical dashed line indicates the maximum likelihood estimate.

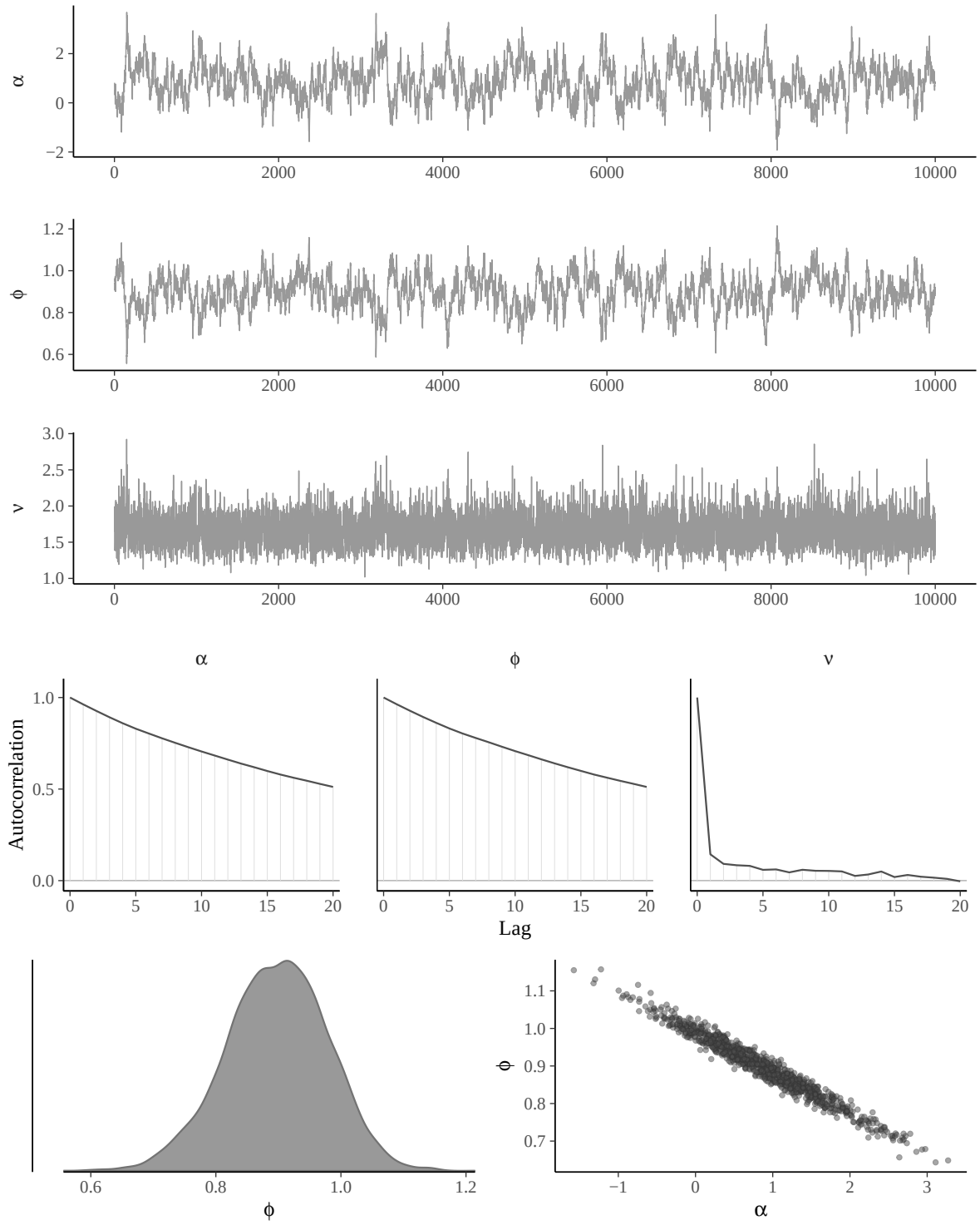


Figure 2: Diagnostics of the Markov chains obtained by running a Gibbs sampler to obtain 10 000 posterior draws from α, ϕ, ν, with trace plots (top), correlograms of individual chains (middle), histogram of the autocorrelation coefficient (bottom left) and scatterplot of pairs (α, ϕ) (bottom right). The estimated effective sample sizes are $\mathsf{ESS}(\alpha) = 226$, $\mathsf{ESS}(\phi) = 217$ and $\mathsf{ESS}(\nu) = 3644$.

2.1 Write down the joint likelihood $L(\phi, \alpha, \nu; \boldsymbol{y}_{1:n})$. [2 pts]

Solution: The joint density can be expressed as the product of conditional densities, using Markov property

$$\begin{aligned} p(\boldsymbol{y}_{1:n} \mid \alpha, \phi, \nu) &= \prod_{t=1}^{n} p(y_t \mid y_{t-1}, \alpha, \phi, \nu) \\ &= (2\pi\nu)^{-n/2} \exp\left[-\frac{1}{2\nu}\sum_{t=1}^{n}\{y_t - \alpha - \phi y_{t-1}\}^2\right] \end{aligned}$$

and treating this expression as a function of (α, ϕ, ν) gives the likelihood $L(\phi, \nu; \boldsymbol{y}_{1:t})$.

2.2 Consider the reference prior [1 pt]

$$p(\alpha, \phi, \nu) \propto \nu^{-1}\, \mathrm{I}(\nu > 0)\mathrm{I}(\alpha \in \mathbb{R})\mathrm{I}(\phi \in \mathbb{R}). \tag{2}$$

Is this prior proper? Justify your answer.

Solution: Since $\upsilon > 0$, the prior is improper because the integral of x^{-1} over the positive half line diverges. It is also improper because $\int_{\mathbb{R}} 1 = \infty$.

2.3 Using the reference prior of Equation (2), derive and provide explicit expressions for the parameters of the posterior distributions [4 pts]

$$\begin{aligned} &p(\alpha \mid \phi, \nu, \boldsymbol{y}_{1:n}) \\ &p(\phi \mid \alpha, \nu, \boldsymbol{y}_{1:n}) \\ &p(\nu \mid \alpha, \phi, \boldsymbol{y}_{1:n}) \end{aligned}$$

Explain how you would use these to obtain posterior draws using a Gibbs sampler.

Solution: Straightforward calculations show that

$$\begin{aligned} p(\alpha \mid \phi, \nu, \boldsymbol{y}_{1:n}) &\sim \mathsf{Gauss}\left\{\frac{\sum_{t=1}^{n}(y_t - \phi y_{t-1})}{n}, \frac{\nu}{n}\right\} \\ p(\phi \mid \alpha, \nu, \boldsymbol{y}_{1:n}) &\sim \mathsf{Gauss}\left\{\frac{\sum_{t=1}^{n}(y_t - \alpha) y_{t-1})}{\sum_{t=1}^{n} y_{t-1}^2}, \frac{\nu}{\sum_{t=1}^{n} y_{t-1}^2}\right\} \\ p(\nu \mid \alpha, \phi, \boldsymbol{y}_{1:n}) &\sim \mathsf{inv.gamma}\left\{\frac{n}{2}, \frac{1}{2}\sum_{t=1}^{n}\left(y_t - \alpha - \phi y_{t-1}\right)^2\right\} \end{aligned}$$

Gibbs sampler: starting from say the maximum likelihood estimator, alternate sampling from each conditional distribution in turn, treating other parameters as fixed.

2.4 Suppose you to select a more informative prior for the variance by taking $\nu \sim \mathsf{inv.gamma}(\beta, \beta)$. Derive the expected value of the $\mathsf{inv.gamma}(\beta, \beta)$ prior and use moment matching to select the value β such that the apriori expected value for ν is 1.5. [2 pts]

Solution: The density of the inverse gamma is

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1}\exp(-\beta/x)\mathrm{I}(x>0)$$

so we have for $\alpha > 1$

$$\begin{aligned}\mathsf{E}(X) &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{-(\alpha-1)-1}\exp(-\beta/x)\mathrm{d}x \\ &= \frac{\beta\Gamma(\alpha-1)}{\Gamma(\alpha)}\int_0^\infty f(x;\alpha-1,\beta)\mathrm{d}x\end{aligned}$$

and noting that $\Gamma(x+1) = x\Gamma(x)$, we recover $\mathsf{E}(X) = \beta/(\alpha-1)$. If we set $\beta/(\beta-1) = 1.5$, then $\beta = 3$.

2.5 Figure 2 on page 3 shows diagnostic plots for the Gibbs sampler. Comment on the stationarity and mixing, and diagnose any problems related to sampling. [2 pts]

Solution: The chains all appear to have reached their stationary distribution. It's obvious that there is a mixing problem for the mean parameters, owing to the strong correlation between (α,ϕ) that leads to slow exploration of the state space and low effective sample sizes.

2.6 Figure 2 shows some posterior draws of ϕ exceed one, which corresponds to AR(1) nonstationary models with infinite variance. How could you enforce stationarity? [1 pt]

Solution: A simple option is to take a truncated prior which has range $[-1,1]$ or assigns little to no prior probability there, e.g., a penalized complexity prior. Another option is to shrink towards zero.

2.7 In line with the previous two subquestions, describe modifications of your algorithm that would improve the efficiency of the sampling of posterior draws. [2 pts]

Solution: This could be remedied by making block updates using Gibbs (this is a linear regression) or via a within Metropolis–Hastings step, or by reparametrizing the model using, e.g., the centered parametrization $\mu + \phi(y_{t-1} - \mu)$.
With a simple redefinition $p(\phi) = \frac{1}{2}\mathrm{I}(|\phi| \leq 1)$, the conditional $p(\phi \mid \alpha, \nu, \boldsymbol{y}_{1:n})$ for the Gibbs sampler is simply a truncated Gaussian, which can be sampled using either accept-reject or inversion sampling. Otherwise, one would need to use Metropolis-within-Gibbs for ϕ.

2.8 For February 2012, the analysis yields a one-step ahead forecast, obtained from the posterior predictive distribution $p(\tilde{y}_{n+1} \mid \boldsymbol{y}_{1:n})$, of 5.39% with an equitailed 50% credible interval of [4.48%, 6.30%]. Explain in your own words how you would calculate such interval based on the posterior draws $(\alpha_{(b)}, \phi_{(b)}, \sigma_{(b)})$ $(b = 1,\ldots,B)$. [2 pts]

Solution: For each of the B draws of the Markov chain, say $(\alpha_{(b)}, \phi_{(b)}, \sigma_{(b)})$ $(b = 1, \ldots, B)$ sample a single realization from the posterior predictive $\mathsf{Gauss}(\alpha_{(b)} + \phi_{(b)} y_n, \sigma^2_{(b)})$.
Compute the empirical quartiles and return these.

Question 3. Betting on outcomes of soccer matches

Let $Y_{ij\mathrm{h}}$ (resp. $Y_{ij\mathrm{v}}$) denote the score of the home (resp. visitor) team for a soccer match opposing teams i and j. Maher (1982) suggests modelling the scores as

$$\begin{aligned} Y_{ij\mathrm{h}} &\sim \mathsf{Poisson}\{\exp(\delta_0 + \delta + \alpha_i + \beta_j)\}, \\ Y_{ij\mathrm{v}} &\sim \mathsf{Poisson}\{\exp(\delta_0 + \alpha_j + \beta_i)\}, \qquad i \neq j; i, j \in \{1, \ldots, 20\}, \end{aligned}$$

where δ_0 is the intercept parameter, α_i represents the offensive strength of the team, β_j the defensive strength of team j and δ is the **home advantage**.

To complete the specification of the model, we take the offensive and defensive strength to be independent and identically distributed $\alpha_i \sim \mathsf{Gauss}(0, \tau_\alpha^{-1})$ and $\beta_j \sim \mathsf{Gauss}(0, \tau_\beta^{-1})$ for $j = 1, \ldots, 20$, along with $\delta \sim \mathsf{Gauss}(0, 4)$ and $p(\delta_0) \propto 1$. We assign penalized-complexity priors to τ_α and τ_β.

The data are the number of goals from matches in the English Football League (EFL) in 2024. Each of the 20 teams in the division plays one match at home and one away against every other team, for a total of 380 matches. The data for those matches before the Championship contain the following variables ($n = 760$ observations in long format):

- `score`: number of goals of `team` during a match
- `team`: categorical variable giving the name of the team that scored the goals
- `opponent`: categorical variable giving the name of the adversary
- `home`: binary variable, 1 if `team` is playing at home, 0 otherwise.

score	team	opponent	home
0	Burnley	Man City	1
3	Man City	Burnley	0
2	Arsenal	Nottingham Forest	1
1	Nottingham Forest	Arsenal	0
1	Bournemouth	West Ham	1

Table 1: First five observations out of $n = 760$ for the EFL matches.

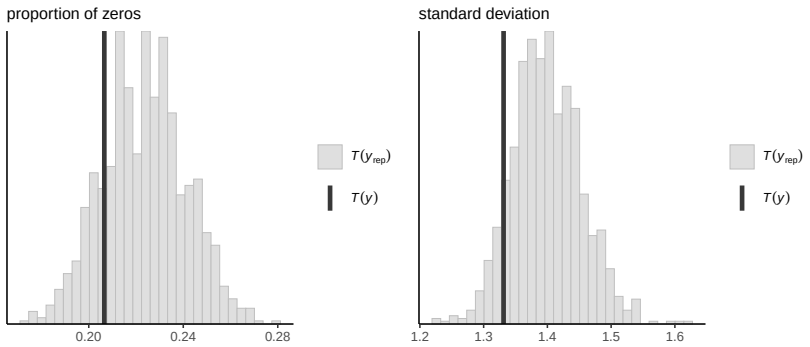


Figure 3: Posterior predictive checks for two test functions $T(y)$: the number of zero goals scored out of 760 scores (left), and the standard deviation of goals (right)

3.1 We run a Markov chain Monte Carlo (MCMC) algorithm for 4000 draws post burn-in to get posterior samples. Describe how you could obtain the following quantities by Monte Carlo methods based on these (recall that δ is the parameter for the home advantage): [3 pts]

i. A 89% equitailed credible interval for the **mean increase** due to the home advantage (*Hint: careful! the Poisson model is fitted with a log link function.*

ii. The posterior probability that the "home advantage" benefits the team playing at home.

> **Solution:** For the first, we simply exponentiate the draws from δ and compute sample quantiles at level 0.055 and 0.945 (i.e., the 220th and 3780th order statistics, $\exp\delta_{(220)}$ and $\exp\delta_{(3780)}$). For the second, we compute the sample proportion of δ that exceeds zero, $B^{-1}\sum_{b=1}^{B}\mathrm{I}(\delta_b > 0)$.

3.2 Of interest are the 3rd to 6th ranked teams, who go to Championship play-offs. The Championship has two rounds: semifinals and finals. During the semifinals, each team plays two matches against each other (with the 3rd ranked team playing against the 6th ranked, and 4th against 5th, starting with the first match in the lower-ranked team home stadium and the second match in the high-ranked home stadium). The team with the higher number of goals in each semifinal moves to the next round. The final between the two finalists takes place in a neutral venue for a single match. See Figure 4 for an example outcome. [3 pts]

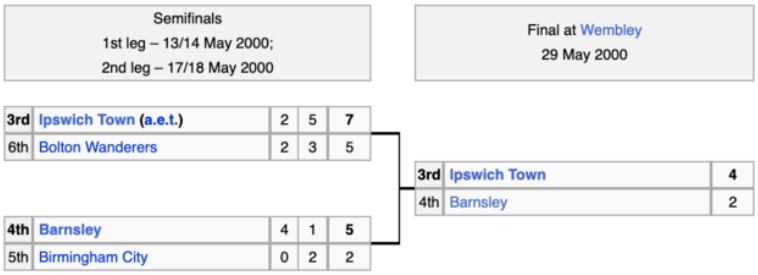


Figure 4: Play-offs structure source: Wikipedia, The Rambling Man - Own work, CC BY-SA 4.0

As part of your EPF league betting pool and knowing the identity of the teams partaking in the Championship, **briefly** sketch how you could get a probabilistic forecast of the winner

of the Championship. *Assume, for simplicity, that the total score in the semifinals, or the final won't feature ties.*

> **Solution:** If we ignore for a minutes ties, we can simply simulate match outcomes from the posterior predictive, adding the home advantage to each team in turn. The players with the highest cumulative scores then advance to the finals, and we simulate from Poisson (without δ factor) based only on the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ factors.

3.3 Figure 3 shows posterior predictive checks for the model. Give a pseudo-code to reproduce the posterior predictive checks of Figure 3, given the sample vector of observations $\boldsymbol{y}$ representing the score and a collection of posterior draws $\boldsymbol{\theta}$ for each model parameter. [2 pts]

> **Solution:** The posterior predictive check consists in sampling from the posterior predictive for each of the 760 combination of covariates from the Poisson model, compute a summary statistic, and repeat this N times. We then compare the rank of the statistic on the original sample $T(\boldsymbol{y})$ with the N copies $T(\boldsymbol{y}^{\text{rep}})$.

3.4 The model is parameterized in terms of the precision $\tau = \sigma^{-2}$ which is assigned a penalized complexity (PC) prior. The PC prior is invariant to reparameterization and that for the scale parameter (std. deviation) $\sigma > 0$ of a Gaussian is exponential with rate λ, meaning it has density $f(\sigma;\lambda) = \lambda\exp(-\lambda\sigma)$ for $\lambda > 0$. What is the density for the PC prior of the precision τ? *Hint: do a change of variable and derive the corresponding density.* [2 pts]

> **Solution:** Write $\sigma = \tau^{-1/2}$, so that $\mathrm{d}\sigma = -\frac{1}{2}\tau^{-3/2}\mathrm{d}\tau$. Then, the PC prior for τ is, using the change of variable formula,
>
> $$f(\tau;\lambda) = \frac{\lambda}{2}\tau^{-3/2}\exp(-\lambda\tau^{-1/2}).$$
>
> This is the density of a Gumbel type-2 distribution.

3.5 Even if $f(x)$ is not normalized (i.e., a positive function such that $\int f(x)\mathrm{d}x = c_f < \infty$), we can still use rejection sampling. Propose a rejection sampling algorithm for the precision $\tau_\alpha \mid \boldsymbol{\alpha}$ if we use the PC prior. *Hint: obtain an upper bound for the unnormalized conditional density of* $\tau_\alpha \mid \cdot$ *by using the fact* $\exp(-x) \leq 1$ *for* $x > 0$. [2 pts]

> **Solution:** With the unnormalized log posterior, keeping only terms that depend on τ, as
>
> $$p(\tau \mid \boldsymbol{\delta}) \propto \tau^{(n-3)/2}\exp\left(-\lambda\tau^{-1/2} - \tau\frac{\sum_{j=1}^J \delta_j^2}{2}\right) \leq \tau^{(n-3)/2+1-1}\exp\left(-\tau\frac{\sum_{j=1}^J \delta_j^2}{2}\right)$$
>
> since $\exp(-x)$ is a decreasing function of x and $x > 0$ implies the function is less than one. We can thus propose from a gamma distribution with shape $(n-3)/2+1$ and rate $0.5\sum_{j=1}^J \delta_j^2$.

3.6 Suppose that we set the parameter of the PC prior for the Gaussian scale such that $\sigma = 1$ corresponds to the 0.99 percentile of the (exponential) PC prior for σ. Report the value of λ that satisfies this constraint. [2 pts]

Solution: The survival function of the exponential distribution is $\exp(-\lambda) = 0.01$, so taking logarithm on both sides gives $\lambda_0 = -\log(0.01)$.