BAYESIAN STATISTICS A. Guglielmi 03.07.2020

Properly justify all your answers. Use the indicator function to denote the support of a distr..

X **Exercise 1** Let $X_1, \dots, X_n$ be an i.i.d. sample from the double exponential distribution (also known as Laplace distribution). This distribution has the following density, w.r.t. the Lebesgue measure on $\mathbb{R}$:

$$f(x; \mu, \lambda) = \frac{1}{2\lambda} \exp\left(-\frac{|x-\mu|}{\lambda}\right), \quad \text{with } \mu \in \mathbb{R}, \lambda > 0 \tag{1}$$

Assume $\mu = 2$ for the rest of the exercise.

1. Derive the likelihood with observed data $\boldsymbol{x} = (x_1, \dots, x_n)$. Then derive a conjugate prior $\pi_1(\lambda)$ for λ, with the parametrization of the prior distribution leading to $\mathrm{E}[\lambda] = \frac{b}{a-1}$, when the expectation exists. Find also the hyperparameters of the posterior distribution $\pi_1(\lambda|\boldsymbol{x})$.

   (**Hint**: *if you are not able to find the conjugate prior for λ with the correct parameterization in one step, you could first find a conjugate prior for $\tau = 1/\lambda$ such that $\mathrm{E}[\lambda] = \mathrm{E}[1/\tau] = \frac{b}{a-1}$ and then consider the transformed distribution for λ as the prior*)

2. A priori belief is expressed in terms of a previously observed sample $\boldsymbol{x}^{old} = (1.7, 1.9, 2.05, 2.15, 2.2)$. Use this *equivalent sample* to fix the hyperparameters of the conjugate prior $\pi_1(\lambda)$.

   (**Hint**: *compute the posterior mean of λ with old data $\boldsymbol{x}^{old}$ and express it as a convex linear combination of ...*)

3. With observed data $\boldsymbol{x}$ such that $n = 75$ and $\sum_{i=1}^n |x_i - \mu| = 19.7$, write down the hyperparameters of the posterior density $\pi_1(\lambda|\boldsymbol{x})$ when the prior hyperparameters are those at point 2. Then find the mean and variance of the posterior distribution $\pi_1(\lambda|\boldsymbol{x})$ of λ. Provide also the posterior 95% HPD (*highest posterior density*) interval for λ, using a suitable approximation.

   (**Hint**: *use asymptotic properties of the posterior under regularity conditions; moreover, remember that* $\displaystyle\int_0^{+\infty} \frac{1}{t^2} \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} dt = \frac{b^2}{(a-1)(a-2)}$, $a > 2$)

4. Find the Jeffreys' prior $\pi_J(\lambda)$ for λ and the corresponding posterior distribution with the data at point 3.

   (**Hint**: *If Z has the double exponential distribution as in* (1), *then* $|Z - \mu| \sim Exp(\lambda^{-1})$, *that is the exponential distribution with mean equal to* λ)

5. With the prior $\pi_1(\lambda)$ at point 2. and data at point 3., find the predictive density for the next observation X_{n+1}. Then evaluate the predictive density at $x_{n+1} = 2.5$.

Exercise 2 Consider the probit regression model

$$\begin{aligned} Y_i|\boldsymbol{x}_i, \boldsymbol{\beta} &\overset{\text{ind.}}{\sim} Be(p_i), \; i = 1, \dots, n \\ p_i &= \Phi(\boldsymbol{x}_i^t \boldsymbol{\beta}) = \Phi(\beta_1 x_{i1} + \dots + \beta_p x_{ip}) \\ \boldsymbol{\beta} &\sim \mathcal{N}_p(\boldsymbol{b}_0, B_0), \end{aligned} \tag{2}$$

where $\boldsymbol{b}_0 \in \mathbb{R}^p$, B_0 is a $p \times p$ covariance matrix and Φ is the standard gaussian d.f.. Introducing suitable latent variables, describe a Gibbs sampler algorithm to simulate from the posterior distribution of $\boldsymbol{\beta}$, given data $y_1, \dots, y_n$, according to model (2).

Solution of Exercise 1

1. The likelihood function is equal to

$$L(\mu, \lambda; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \mu, \lambda) = \prod_{i=1}^{n} \frac{1}{2\lambda} \exp\left(-\frac{|x_i - \mu|}{\lambda}\right) = \frac{1}{2^n \lambda^n} \exp\left(-\frac{1}{\lambda}\sum_{i=1}^{n}|x_i - \mu|\right), \lambda > 0.$$

Assuming $\mu = 2$, it is clear that the conjugate prior for λ is the inverse-gamma distribution, $\lambda \sim inv - gamma(a, b)$, with density

$$\pi_1(\lambda) = \frac{b^a}{\Gamma(a)} \frac{1}{\lambda^{a+1}} e^{-\beta/\lambda} \mathbf{1}_{(0,+\infty)}(\lambda)$$

so that $\mathrm{E}[\lambda] = \frac{b}{a-1}$ when $b > 0$ and $a > 1$. The posterior distribution is obtained via Bayes' theorem as

$$\pi(\lambda \mid \boldsymbol{x}) \propto \pi(\lambda) L(2, \lambda; \boldsymbol{x}) \propto \frac{1}{\lambda^{a+n+1}} \exp\left(-\frac{1}{\lambda}\left[b + \sum_{i=1}^{n}|x_i - \mu|\right]\right) \mathbf{1}_{(0,+\infty)}(\lambda)$$

that is $\lambda|\boldsymbol{x} \sim inv - gamma(a_n, b_n)$. The hyperparameters of the posterior distribution are $a_n = a + n$ and $b_n = b + \sum_{i=1}^{n}|x_i - \mu|$.

2. The posterior mean of λ given the old dataset $\boldsymbol{x}^{old}$ with sample size $m = 5$ is

$$\mathrm{E}[\lambda \mid \boldsymbol{x}^{old}, \mu = 2] = \frac{b + \sum_{i=1}^{5}|x_i^{old} - 2|}{a + m - 1} = \frac{a-1}{a+m-1}\frac{b}{a-1} + \frac{m}{a+m-1}\frac{1}{m}\sum_{i=1}^{5}|x_i^{old} - 2|$$

and then we have $a - 1 = m = 5$ and $b = \sum_{i=1}^{5}|x_i^{old} - 2| = 0.8$. Hence $a = 6$, $b = 0.8$.

3. First of all, observe that, given the data, $a_n = a + n = 81$ and $b_n = b + \sum_{i=1}^{n}|x_i - \mu| = 20.5$, so that $\lambda|\boldsymbol{x} \sim inv - gamma(81, 20.5)$.

   Remember that the variance of the $inv - gamma(a, b)$ distribution is

$$\frac{b^2}{(a-1)(a-2)} - \frac{b^2}{(a-1)^2} = \frac{b^2}{(a-1)^2(a-2)}$$

   The posterior mean and variance of λ are equal to

$$\mathrm{E}[\lambda \mid \boldsymbol{x}] = \frac{20.5}{80} = 0.256 \qquad Var(\lambda \mid \boldsymbol{x}) = \frac{(20.5)^2}{80^2 \times 79} = 0.00083$$

   and then $\lambda \mid \boldsymbol{x} \sim inv - gamma(81, 20.5) \simeq N(0.256, 0.00083)$. The 95% HPD for λ is approximately given by

$$CI_{0.95} = \mathrm{E}[\lambda \mid \boldsymbol{x}] \pm z_{0.975}\sqrt{Var(\lambda \mid \boldsymbol{x})} = 0.256 \pm z_{0.975} \times 0.0288$$

   with $z_{0.975}$ denotes the quantile of order 0.975 of the standard Gaussian distribution, then the credible interval is

$$(0.200; 0.312)$$

4. The Jeffreys' prior $\pi_J(\lambda)$ is proportional to $\sqrt{|I(\lambda)|}$, where $I(\lambda)$ denotes Fisher's information. Then we have, assuming μ known,

$$
\begin{aligned}
I(\lambda) &= \mathrm{E}\left[-\frac{d^2}{d\lambda^2}\log f(X_1;2,\lambda))\right] = \mathrm{E}\left[-\frac{d^2}{d\lambda^2}\left(-\frac{1}{\lambda}|X_1-\mu| - \log(\lambda)\right)\right] \\
&= \mathrm{E}\left[\frac{2}{\lambda^3}|X_1-\mu| - \frac{1}{\lambda^2}\right] = \frac{2}{\lambda^3}\,\mathrm{E}\left[|X_1-\mu|\right] - \frac{1}{\lambda^2} \\
&= \frac{2}{\lambda^3}\lambda - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}
$$

Then the Jeffreys' prior, proportional to $\sqrt{|I(\lambda)|}$, is given by:

$$\pi_J(\lambda) \propto \sqrt{\frac{1}{\lambda^2}}\mathbf{1}_{(0,+\infty)}(\lambda) = \frac{1}{\lambda}\mathbf{1}_{(0,+\infty)}(\lambda)$$

and the corresponding posterior distribution is

$$\pi_J(\lambda \mid \boldsymbol{x}) \propto \pi_J(\lambda)L(\lambda;\boldsymbol{x},\mu) = \frac{1}{2^n\lambda^{n+1}}\exp\left(-\frac{1}{\lambda}\sum_{i=1}^{n}|x_i-\mu|\right)\mathbf{1}_{(0,+\infty)}(\lambda)$$

which identifies the kernel of an inverse gamma distribution.
Hence $\pi_J(\lambda \mid \boldsymbol{x})$ is $inv-gamma(n, \sum_{i=1}^{n}|x_i-\mu|)$, specifically $\lambda \mid \boldsymbol{x} \sim inv-gamma(75, 19.7)$.

5. The predictive density function can be derived as

$$
\begin{aligned}
m_{X_{n+1}|\mathbf{X}}(x_{n+1} \mid \boldsymbol{x}) &= \int_0^{+\infty} f(x_{n+1} \mid \lambda)\pi(\lambda \mid \boldsymbol{x})d\lambda \\
&= \int_0^{+\infty} \frac{1}{2\lambda}\exp\left(-\frac{1}{\lambda}|x_{n+1}-\mu|\right)\frac{(b_n)^{a_n}}{\Gamma(a_n)}\frac{1}{\lambda^{a_n+1}}\exp\left(-\frac{b_n}{\lambda}\right)d\lambda \\
&= \frac{1}{2}\frac{(b_n)^{a_n}}{\Gamma(a_n)}\frac{\Gamma(a_n+1)}{(b_n+|x_{n+1}-\mu|)^{a_n+1}}\int_0^{+\infty}\frac{(b_n+|x_{n+1}-\mu|)^{a_n+1}}{\Gamma(a_n+1)}\frac{1}{\lambda^{a_n+2}}\exp\left(-\frac{b_n+|x_{n+1}-\mu|}{\lambda}\right)d\lambda \\
&= \frac{1}{2}\frac{(b_n)^{a_n}}{\Gamma(a_n)}\frac{\Gamma(a_n+1)}{(b_n+|x_{n+1}-\mu|)^{a_n+1}}
\end{aligned}
$$

Since we have $a_n = 81$ and $b_n = 20.5$, the predictive density assumes value

$$m_{X_{n+1}|\mathbf{X}}(x_{n+1} \mid \boldsymbol{x}) = \frac{1}{2}\frac{(20.5)^{81}}{\Gamma(81)}\frac{\Gamma(82)}{(20.5+|x_{n+1}-\mu|)^{82}} = \frac{81}{2}\frac{(20.5)^{81}}{(20.5+|x_{n+1}-\mu|)^{82}}$$

and

$$m_{X_{n+1}|\mathbf{X}}(2.5 \mid \boldsymbol{x}) = \frac{81}{2}\frac{(20.5)^{81}}{(21)^{82}} = 0.274$$

∎

Solution of Exercise 2
We introduce latent variables $Z_1, \ldots, Z_n$ such that

$$Y_i|\boldsymbol{x}_i = \begin{cases}1 & \text{if } Z_i > 0 \\ 0 & \text{otherwise}\end{cases}, \; i = 1, \ldots, n \tag{3}$$

$$Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{x}_i^t\boldsymbol{\beta}, 1). \tag{4}$$

It is straightforward to check that the likelihood in (2) is equivalent to (3)-(4). The *parameter* is now $(\boldsymbol{\beta}, \boldsymbol{Z})$, where $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^t$. We build a Gibbs sampler to simulate from the joint posterior law $\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{Z}|y_1, \ldots, y_n)$, but in the end we will be interested only in the marginal posterior $\mathcal{L}(\boldsymbol{\beta}|y_1, \ldots, y_n)$. In order to derive a Gibbs sampler, we need the full-conditionals, that are proportional to the joint law of data and the parameter vector:

$$\mathcal{L}(Y_1, \ldots, Y_n, \boldsymbol{\beta}, \boldsymbol{Z}) \propto \mathcal{L}(Y_1, \ldots, Y_n|\boldsymbol{\beta}, \boldsymbol{Z}) \times \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{Z}) = \mathcal{L}(Y_1, \ldots, Y_n|\boldsymbol{Z}) \times \mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}) \times \pi(\boldsymbol{\beta}).$$

Then, the full-conditionals are:

- $\mathcal{L}(\boldsymbol{\beta}|\boldsymbol{Z}, \boldsymbol{Y}) \propto \mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}) \times \pi(\boldsymbol{\beta})$. This last distribution is the posterior of a linear model with Gaussian likelihood, where the *data* are $Z_1, \ldots, Z_n$, i.e.

$$\mathcal{L}(\boldsymbol{\beta}|\boldsymbol{Z}, \boldsymbol{Y}) = \mathcal{N}_p(\boldsymbol{b}_n, B_n), \quad B_n = (X^t X + B_0^{-1})^{-1}, \ \boldsymbol{b}_n = B_n(X^t X \hat{\boldsymbol{\beta}} + B_0^{-1} \boldsymbol{b}_0).$$

  Here X is the design matrix with $\boldsymbol{x}_i$'s as rows, and $\hat{\boldsymbol{\beta}} = (X^t X)^{-1} X^t \boldsymbol{Z}$.

- $\mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{Y}) \propto \mathcal{L}(Y_1, \ldots, Y_n|\boldsymbol{\beta}, \boldsymbol{Z}) \times \mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ (\mathbf{1}(y_i = 1)\mathbf{1}(Z_i > 0) + \mathbf{1}(y_i = 0)\mathbf{1}(Z_i < 0)) \, \phi(Z_i; \boldsymbol{x}_i^t \boldsymbol{\beta}, 1) \right\}$, i.e. $Z_1, \ldots, Z_n$ are independent given $\boldsymbol{\beta}, \boldsymbol{y}$, and

$$Z_i|\boldsymbol{\beta}, \boldsymbol{y} \sim \begin{cases} \mathbf{1}(Z_i > 0)\phi(Z_i; \boldsymbol{x}_i^t \boldsymbol{\beta}, 1) & \text{if } y_i = 1 \\ \mathbf{1}(Z_i < 0)\phi(Z_i; \boldsymbol{x}_i^t \boldsymbol{\beta}, 1) & \text{if } y_i = 0. \end{cases}$$

  The last two expressions denote the truncated Gaussian densities with support $(0, +\infty)$ and $(-\infty, 0)$, respectively.

∎

BAYESIAN STATISTICS A. Guglielmi × 16.06.2020

Properly justify all your answers. Use the indicator function to denote the support of a distr..

× **Exercise 1** Let $X_1, \ldots, X_n$, conditionally to θ, be a sample from the absolutely continuous uniform distribution on the interval $[0, \theta]$, with $\theta > 0$, with density

$$f(x;\theta) = \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x), \qquad \theta > 0 \tag{1}$$

Given the observed sample, we are interested in making inference on the parameter θ.

1. Derive the likelihood with observed data $(x_1, \ldots, x_n)$, with $x_i > 0$ for all $i = 1, \ldots, n$. Then show that the Pareto density (w.r.t. the Lebesgue measure)

$$\pi(\theta; m, k) = \frac{km^k}{\theta^{k+1}}\mathbf{1}_{[m,+\infty)}(\theta), \qquad m > 0,\ k > 0 \tag{2}$$

is conjugate to the likelihood. Find also the hyperparameters in the posterior distribution.

(**Hint**: *Recall that* $\mathbf{1}_{[0,\theta]}(x) = \mathbf{1}_{[x,+\infty)}(\theta)$, *so that* $\prod_{i=1}^n \mathbf{1}_{[0,\theta]}(x_i) = \mathbf{1}_A(\theta)$ *where A is ....*)

2. Given the observed sample $\boldsymbol{x} = (4.2, 5.6, 6.1, 2.5, 3.4)$ of size $n = 5$, fix $m = 3$ and $k = 5$ in (2). Write down the posterior hyperparameters. Then provide the posterior point estimate of θ under the quadratic loss function.

3. Given the observed sample at point 2., test the hypotheses $\{H_0 : \theta = 7\}$ versus $\{H_1 : \theta \neq 7\}$, when the prior density under H_1 is the conjugate prior derived at point 2., by computing the Bayes factor.

4. Assuming the conjugate prior derived at point 2., compute the predictive probability $\mathbb{P}[X_6 > 6.5 \mid \boldsymbol{x}]$ that the next observation is larger than 6.5.

(**Hint**: *Compute first* $\mathbb{P}[X_6 > 6.5 \mid \theta]$)

× **Exercise 2** Let $X_1, \ldots, X_n$ be a sample from the Bernoulli distribution, i.e.

$$X_1, \ldots, X_n | \theta \overset{\text{i.i.d.}}{\sim} Be(\theta), \quad \theta \in (0,1),$$

that is every X_i is equal to 1 with probability θ and equal to 0 with probability $1 - \theta$. We consider a discrete prior distribution for θ, restricting the support to $\Theta^* = \{0.2, 0.4, 0.6, 0.8\}$, and with prior probabilities given in the following table

θ	0.2	0.4	0.6	0.8
$\pi(\theta)$	0.1	0.2	0.4	0.3

Table 1: Discrete prior distribution $\pi(\theta)$.

1. Derive the likelihood with $n = 15$ observed data $\boldsymbol{x} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1)$; then compute the posterior distribution of the parameter θ, given the prior distribution specified in Table 1.

(**Hint**: *Recall that θ is discrete, so that the posterior distribution will be discrete as well. It is sufficient to compute via Bayes' theorem $\pi(\theta|\boldsymbol{x})$ for any $\theta \in \Theta^*$. Pay attention to the numerical evaluation of these values.*)

2. Provide the posterior estimates of θ, under the quadratic loss function and the $0-1$ loss function.

3. Assuming the prior distribution for θ in Table 1, test the hypotheses $H_0 : \theta \in \{0.2, 0.4\}$ versus $H_1 : \theta \in \{0.6, 0.8\}$ by computing the Bayes factor, given the observed sample introduced at point 1.

   (**Hint**: *in this case, compute the Bayes factor as the ratio of the posterior odds versus the prior odds.*)

4. Compute the predictive distribution of X_{16} given $\boldsymbol{x}$, assuming the prior distribution in Table 1 with observed data at point 1.

Solution of **Exercize 1**.

1. We have that the likelihood function is equal to

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbf{1}_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \mathbf{1}_{[\max(\boldsymbol{x}),+\infty)}(\theta).$$

   Then given the prior in (2), the posterior becomes

$$\pi(\theta \mid \boldsymbol{x}) \propto \pi(\theta) L(\theta; \boldsymbol{x}) = \frac{km^k}{\theta^{k+1}} \mathbf{1}_{[m,+\infty)}(\theta) \frac{1}{\theta^n} \mathbf{1}_{[\max(\boldsymbol{x}),+\infty)}(\theta) \propto \frac{1}{\theta^{k+n+1}} \mathbf{1}_{[\max(m,\boldsymbol{x}),+\infty)}(\theta)$$

   which is the kernel of the Pareto distribution with parameters $m_1 = max(m, \max_i x_i)$ and $k_1 = k + n$.

2. With data $\boldsymbol{x} = (4.2, 5.6, 6.1, 2.5, 3.4)$, $\max_i x_i = 6.1$, $m = 3$, $k = 5$, so that $m_1 = max(m, \max_i x_i) = 6.1$ and $k_1 = k + n = 10$. Hence the posterior distribution of theta is the Pareto density $\pi(\theta; m_1, k_1)$.

   We know that, under the quadratic loss function, the posterior point estimation is equal to the posterior expected value. Then we have

$$\begin{aligned} \mathrm{E}[\theta \mid \boldsymbol{x}] &= \int_{\mathbb{R}} \theta \frac{k_1 m_1^{k_1}}{\theta^{k_1+1}} \mathbf{1}_{[m_1,+\infty)}(\theta) d\theta = k_1 m_1^{k_1} \int_{m_1}^{+\infty} \theta^{-k_1} d\theta = \frac{k_1 m_1^{k_1}}{k_1 - 1} \frac{1}{m_1^{k_1-1}} \\ &= \frac{k_1}{k_1 - 1} m_1 = \frac{10}{9} \times 6.1 \simeq 6.7778. \end{aligned}$$

3. We want to test the following hypotheses

$$H_0 : \theta = 7 \text{ vs } H_1 : \theta \neq 7$$

   that is a point hypothesis H_0 against a "diffuse" alternative H_1. The corresponding Bayes factor is

$$BF_{01}(\boldsymbol{x}) = \frac{\prod_{i=1}^{n} f(x_i; \theta = 7)}{m_1(\boldsymbol{x})},$$

   where

$$\begin{aligned} m_1(\boldsymbol{x}) &= \int_{\mathbb{R}} \prod_{i=1}^{n} f(x_i; \theta) \pi(\theta) d\theta = \int_{\mathbb{R}} \frac{1}{\theta^n} \mathbf{1}_{[\max(\boldsymbol{x}),+\infty)}(\theta) \frac{km^k}{\theta^{k+1}} \mathbf{1}_{[m,+\infty)}(\theta) d\theta \\ &= km^k \int_{m_1}^{+\infty} \frac{1}{\theta^{k_1+1}} d\theta = \frac{km^k}{k_1} \left[\frac{1}{\theta^{k_1}} \right]_{+\infty}^{m_1} = \frac{km^k}{k_1 m_1^{k_1}} = \frac{5}{10} \frac{3^5}{(6.1)^{10}} \end{aligned}$$

while the numerator of the BF is equal to $\prod_{i=1}^{n} f(x_i; \theta = 7) = 1/7^5 \mathbf{1}_{[6.1,\infty)}(7)$, so that

$$BF_{01}(\boldsymbol{x}) = \frac{\frac{1}{7^5}}{\frac{5}{10}\frac{3^5}{(6.1)^{10}}} = 2\frac{(6.1)^{10}}{(21)^5} \simeq 34.9327$$

and $\log BF_{01} \simeq 7.1068$. There is strong evidence in favour of H_0.

4. We have that

$$\mathbb{P}[X_6 > 6.5 \mid \boldsymbol{x}] = \int_{\mathbb{R}} \mathbb{P}[X_6 > 6.5 \mid \theta]\pi(\theta \mid \boldsymbol{x})d\theta.$$

By (1), $\mathbb{P}[X_6 > 6.5 \mid \theta] = 1 - F_{X_6}(6.5|\theta) = \left(1 - \frac{6.5}{\theta}\right)\mathbf{1}_{[6.5,+\infty)}(\theta)$, so that

$$\begin{aligned}
\mathbb{P}[X_6 > 6.5 \mid \boldsymbol{x}] &= \int_{\mathbb{R}} \mathbb{P}[X_6 > 6.5 \mid \theta]\pi(\theta \mid \boldsymbol{x})d\theta = \int_{\mathbb{R}} \left(1 - \frac{6.5}{\theta}\right)\frac{k_1 m_1^{k_1}}{\theta^{k_1+1}}\mathbf{1}_{[6.5,+\infty)}(\theta)d\theta \\
&= k_1 m_1^{k_1}\left(\frac{1}{k_1}\left[\frac{1}{\theta^{k_1}}\right]_{+\infty}^{6.5} - \frac{6.5}{k_1+1}\left[\frac{1}{\theta^{k_1+1}}\right]_{+\infty}^{6.5}\right) = 10(6.1)^{10}\left(\frac{1}{10(6.5)^{10}} - \frac{6.5}{11}\frac{1}{(6.5)^{11}}\right) \\
&= \frac{1}{11}\left(\frac{6.1}{6.5}\right)^{10} \simeq 0.0482
\end{aligned}$$

∎

Solution of **Exercize 2**.

1. If $\boldsymbol{x} = (x_1, \dots, x_n) \in \{0,1\}^n$ is a sample from a Bernoulli distribution, then the likelihood is given by

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\mathbf{1}_{(0,1)}(\theta) = \theta^m(1-\theta)^{n-m}\mathbf{1}_{(0,1)}(\theta)$$

with $m = \sum_{i=1}^{n} x_i$. With our data $n = 15$, $m = 6$ and $n - m = 9$. The posterior distribution is discrete as the prior, with the same support, so that

$$\pi(\theta \mid \boldsymbol{x}) = \frac{\pi(\theta)L(\theta;\boldsymbol{x})}{\pi(0.2)L(0.2;\boldsymbol{x}) + \pi(0.4)L(0.4;\boldsymbol{x}) + \pi(0.6)L(0.6;\boldsymbol{x}) + \pi(0.8)L(0.8;\boldsymbol{x})}, \quad \theta = 0.2, 0.4, 0.6, 0.8.$$

Given the prior in Table 1, we have:

θ	$\pi(\theta)$	$\pi(\theta)L(\theta;\boldsymbol{x})$	$\pi(\theta \mid \boldsymbol{x})$
0.2	0.1	8.5900×10^{-7}	0.0612
0.4	0.2	8.2556×10^{-6}	0.5887
0.6	0.4	4.8922×10^{-6}	0.3483
0.8	0.3	4.0300×10^{-8}	0.0028

Table 2: Discrete posterior distribution $\pi(\theta|\boldsymbol{x})$

being the denominator in the Bayes's formula above equal to 1.40471×10^{-5}. Note that the posterior probabilities have been rounded such that $\sum_{i=1}^{4} \pi(\theta_i \mid \boldsymbol{x}) = 1$.

2. Under the quadratic loss function, the posterior point estimation corresponds to the expectation of the posterior distribution:

$$\hat{\theta}_Q = \mathrm{E}[\theta \mid \boldsymbol{x}] = \sum_{i=1}^{4} \theta_i \pi(\theta_i \mid \boldsymbol{x}) = 0.2 \times 0.0612 + 0.4 \times 0.5887 + 0.6 \times 0.3483 + 0.8 \times 0.0028 = 0.45894$$

while under the $0-1$ loss function, the posterior point estimation corresponds to the mode of the posterior distribution, so that

$$\hat{\theta}_{0-1} = \underset{\theta}{\operatorname{argmax}}\, \pi(\theta \mid \boldsymbol{x}) = 0.4$$

3. We want to test

$$H_0 : \theta \in \{0.2, 0.4\} \quad vs \quad H_1 : \theta \in \{0.6, 0.8\}$$

with the prior specified in Table 1. We have computed the posterior with observed data at point 1. The Bayes factor is

$$BF_{01} = \frac{\textit{posterior odds}}{\textit{prior odds}} = \frac{P[\theta \in \{0.2, 0.4\} \mid \boldsymbol{x}]/P[\theta \in \{0.6, 0.8\} \mid \boldsymbol{x}]}{P[\theta \in \{0.2, 0.4\}]/P[\theta \in \{0.6, 0.8\}]}$$

We have that

$$P[\theta \in \{0.2, 0.4\} \mid \boldsymbol{x}] = \pi(\theta = 0.2 \mid \boldsymbol{x}) + \pi(\theta = 0.4 \mid \boldsymbol{x}) = 0.0612 + 0.5877 = 0.6489$$

and

$$P[\theta \in \{0.6, 0.8\} \mid \boldsymbol{x}] = \pi(\theta = 0.6 \mid \boldsymbol{x}) + \pi(\theta = 0.8 \mid \boldsymbol{x}) = 0.3483 + 0.0028 = 0.3511.$$

Similarly we have $P[\theta \in \{0.2, 0.4\}] = 0.3$ and $P[\theta \in \{0.6, 0.8\}] = 0.7$, and the Bayes factor is equal to

$$BF_{01} = \frac{\textit{posterior odds}}{\textit{prior odds}} = \frac{0.6499/0.3511}{0.3/0.7} = 4.3124$$

and $2 \log BF_{01} = 2.922989$. There is weak evidence in favour of the hypothesis H_0, i.e. in favour of θ being equal to 0.2 or to 0.4, instead of being equal to 0.6 or 0.8.

4. We focus on the predictive distribution of the $(n+1)$-th observation, given the observed n-sized sample $\boldsymbol{x}$. Since X_{n+1} is binary, it is sufficient to compute $P(X_{n+1} = 1 \mid \boldsymbol{x})$:

$$P(X_{n+1} = 1 \mid \boldsymbol{x}) = \sum_{\theta} P(X_{n+1} = 1 \mid \theta)\pi(\theta \mid \boldsymbol{x}) = \sum_{\theta} \theta\pi(\theta \mid \boldsymbol{x}) = \mathrm{E}[\theta \mid \boldsymbol{x}] = 0.4589.$$

Hence the predictive probability that $X_{n+1} = 1$ is equal to the posterior mean of θ, because $P(X_{n+1} = 1 \mid \theta) = \theta$. In summary, we have that the predictive distribution of X_{16} is as follows:

$$X_{16} = \begin{cases} 1 \text{ with probability } 0.4589 \\ 0 \text{ with probability } 0.5411 \end{cases} .$$

BAYESIAN STATISTICS A. Guglielmi 04.02.2020

Properly justify all your answers. Use the indicator function to denote the support of a distr..

✗ **Exercise 1** Let us consider the conditional density (w.r.t. the Lebesgue measure)

$$f_X(x;\lambda) = \frac{2^\lambda \lambda}{x^{\lambda+1}} \mathbf{1}_{[2,\infty)}(x), \qquad \lambda > 0 \tag{1}$$

We have observed $n+m$ time-to-events of different subjects (in some time unit), but only n of them were totally observed, while the remaining m are right-censored, that is our data is $\boldsymbol{x} = (x_1, \dots, x_n, x_1^*, \dots, x_m^*)$ with indicator for non-censoring $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n, \delta_{n+1} \dots, \delta_{n+m}) = (1, \dots, 1, 0, \dots, 0)$. We assume that the conditional distribution for our data is (1) and we want to make inference on the unknown parameter λ.

1. Compute the survival function corresponding to density (1).
2. Derive the likelihood with observed data $(\boldsymbol{x}, \boldsymbol{\delta})$. Then find a conjugate prior $\pi(\lambda)$ to the likelihood. Find also the hyperparameters in the posterior distribution.

   (**Hint**: when expressing the likelihood in an "useful" form, remember that $(aa/bb)^{cc} = e^{-cc\log(bb/aa)}$, $aa, bb > 0$).
3. We have previously recorded 5 time-to-events in a similar experiment, $\boldsymbol{z} = (4.2, 4.6, 5.1, 5.2, 5.8)$, all without censoring. Assuming the conjugate prior derived at point 2., fix hyperparameter a equal to the number of observations in $\boldsymbol{z}$, and find the value of hyperparameter b which maximize the marginal density of the past data $\boldsymbol{z}$.
4. The observed sample that we have presently recorded contains $n = 84$ observed time-to-events, such that $\sum_{i=1}^n \log(x_i) = 110.355$, and $m = 16$ right-censored observations, such that $\sum_{i=1}^n \log(x_i^*) = 20.371$, with all x_is and x_j^*s larger or equal than 2. Write down the posterior distribution (i.e. the posterior hyperparameters) and the posterior estimates of λ under the quadratic loss function.
5. Find the Jeffreys prior π_J, and the corresponding posterior distribution, with data at point 4.
6. Given data at point 4. and the conjugate prior π at point 3., test the hypoteses $H_0 : \lambda \leq 1.6$ versus $H_1 : \lambda > 1.6$, by computing the Bayes factor.

   (**Hint**: remember the closeness property of the gamma distributions family wrt the multiplication by a positive constant, that $gamma(aa/2, 1/2) = \chi^2_{aa}$, $\mathbb{P}[\chi^2_{10} \leq 14.5] = 0.849$, and that $\Phi(1.768) = 0.961$.)
7. Based on the conjugate prior specified at point 3., with available data at point 4., find the predictive probability that the time-to-event of a new subject is larger than 5.

   (**Hint**: it should be useful to make the numeric computation on the log scale, as $\frac{aa}{bb} = e^{log(aa) - log(bb)}$.)

Exercise 2 During an industrial process, a stick of length 1 is broken into N bits, with N known. Let $(P_1, \cdots, P_N)$ be the sequence describing the length of the N bits the stick has been broken into. Its distribution is described as follows:
let $V_1, \cdots, V_{N-1} \overset{i.i.d.}{\sim} beta(1, \alpha)$, with $\alpha > 0$ and $V_N = 1$ a.s.; then

$$P_1 := V_1 \qquad P_k := V_k \prod_{i=1}^{k-1}(1 - V_i),\ k = 2, \cdots, N\ . \tag{2}$$

1. Show that, **for any** $k = 2, \ldots, N$, we have

$$\prod_{j=1}^{k}(1 - V_j) = \prod_{j=1}^{k-1}(1 - V_j) - P_k. \tag{3}$$

   Use formula (3) to show that $\sum_{k=1}^{N} P_k = 1$ almost surely.

2. Show that the sequence $\{\mathrm{E}(P_k), k = 1, \ldots, N-1\}$ is decreasing. In particular, show that, for $k = 1, \ldots, N-2$, $\mathrm{E}(P_{k+1}) = \eta(\alpha)\,\mathrm{E}(P_k)$: write down the analytic expression of $\eta(\alpha)$, plot its graph, discuss how α is related to the sequence $P_1, \ldots, P_{N-1}$, check the value of $\eta(1)$ and comment.

Solution of **Exercize 1**.

1. Of course, if $t < 2$, then $S(t; \lambda) = 1$; if $t \geq 2$, then

$$S(t;\lambda) = \int_t^{+\infty} \frac{\lambda 2^\lambda}{x^{\lambda+1}} dx = \lambda 2^\lambda \left| -\frac{x^{-\lambda}}{\lambda} \right|_t^{+\infty} = \frac{2^\lambda}{t^\lambda} \tag{4}$$

2. If there is at least one x_i or x_j^* smaller than 2, the likelihood is equal to 0; otherwise, when $x_1, \ldots, x_n, x_1^*, \ldots, x_m^* \geq 2$, then the likelihood function, for $\lambda > 0$, is equal to

$$\begin{aligned}
L(\lambda; \boldsymbol{x}, \boldsymbol{\delta}) &= \prod_{i=1}^{n} f(x_i;\lambda) \prod_{i=1}^{m} S(x_i^*;\lambda) = \prod_{i=1}^{n} \frac{\lambda 2^\lambda}{x_i^{\lambda+1}} \prod_{i=1}^{m} \frac{2^\lambda}{x_i^{*\lambda}} \\
&= \frac{1}{\prod_{i=1}^{n} x_i} \times \lambda^n 2^{(n+m)\lambda} \frac{1}{(\prod_{i=1}^{n} x_i)^\lambda} \frac{1}{(\prod_{i=1}^{m} x_i^*)^\lambda} \\
&\propto \lambda^n \mathrm{e}^{(n+m)\lambda \log 2} \mathrm{e}^{-\lambda(\log(\prod_{i=1}^{n} x_i \prod_{i=1}^{m} x_i^*))} \\
&= \lambda^n \mathrm{e}^{-\lambda[\sum_1^n \log(x_i/2) + \sum_1^m \log(x_i^*/2)]}
\end{aligned}$$

   Of course, if $\lambda \leq 0$ the likelihood is equal to 0.

   It is clear that a conjugate prior is the gamma distribution, $\lambda \sim gamma(a, b)$, with $a, b > 0$, such that the posterior distribution become

$$\begin{aligned}
\pi(\lambda \mid \boldsymbol{x}, \boldsymbol{\delta}) &\propto \pi(\lambda) \mathrm{L}(\lambda; \boldsymbol{x}, \boldsymbol{\delta}) \propto \lambda^{a-1} \mathrm{e}^{-\lambda b} \lambda^n \mathrm{e}^{-\lambda\left[\sum_{i=1}^{n} \log \frac{x_i}{2} + \sum_{i=1}^{m} \log \frac{x_i^*}{2}\right]} \mathbf{1}_{(0,+\infty)}(\lambda) \\
&= \lambda^{a+n-1} \mathrm{e}^{-\lambda\left[b + \sum_{i=1}^{n} \log \frac{x_i}{2} + \sum_{i=1}^{m} \log \frac{x_i^*}{2}\right]} \mathbf{1}_{(0,+\infty)}(\lambda).
\end{aligned}$$

   Hence, a posteriori λ is distributed as $gamma(a + n, b + S_x + S_x^*)$ distribution, where $S_x = \sum_{i=1}^{n} \log \frac{x_i}{2}$ and $S_x^* = \sum_{i=1}^{m} \log \frac{x_i^*}{2}$.

Exponential family: $f(x|\theta) = h(x)\, e^{\theta^T T(x) - \psi(\theta)}$ (1)

Conjugate density for θ: $\pi(\theta; \mu, \eta_0) = K(\mu, \eta_0)\, e^{\theta^T \mu - \eta_0 \psi(\theta)}$ (2)

1. Consider $x_1, .., x_n$ from (1). Compute the posterior $\pi(\theta | x_1, .., x_n)$. Moreover verify that prior (2) is conjugate to (1) and compute the hyperparameters of the posterior density (i.e. how μ and η_0 update)

$$\pi(\theta|\underline{x}) \propto \left[\left(\prod_{i=1}^n h(x_i)\right) e^{\theta^T \sum_{i=1}^n T(x_i) - n\psi(\theta)}\right]\left[K(\mu, \eta_0)\, e^{\theta^T\mu - \eta_0 \psi(\theta)}\right]$$

$$\propto e^{\theta^T(\sum_{i=1}^n T(x_i) + \mu) - \psi(\theta)(\eta_0 + n)}$$

$$\Longrightarrow \pi(\theta | x_1, .., x_n) = \pi(\theta; \mu^{post}, \eta^{post}) : \begin{cases} \mu^{post} = \mu + \sum_{i=1}^n T(x_i) \\ \eta^{post} = \eta_0 + n \end{cases}$$

This also means that prior and likelihood are conjugate.

2. Instead of (1) consider the multinomial density, where, conditionally to the parameters, each X_i is a K-dimensional vector that sums up to n.
Its conditional density is:

$$f((x_1, .., x_k) | (p_1, .., p_k)) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \qquad \begin{cases} x_1 + \cdots + x_k = n \\ p_1 + \cdots + p_k = 1 \end{cases} \qquad (3)$$

Show that (3) belongs to the exp. family and find $\underline{\theta}, h(\cdot), T(\cdot), \psi(\cdot)$.
Find the conjugate prior for $\underline{\theta}$ using (2).

$$L(\underline{x}, \underline{p}) \propto \frac{n!}{x_1! \cdots x_k!} e^{x_1 \log(p_1) + \cdots + x_k \log(p_k)} = \frac{n!}{x_1! \cdots x_k!} e^{\log(\underline{p})^T \underline{x}}\, \mathbb{1}_{supp}(\underline{x})$$

$supp(\underline{x}) = \{\underline{x} : x_i = 0, 1, 2, .., \ x_1 + \cdots + x_k = n\}$

$\underline{\theta} = \log(\underline{p})$

$h(\cdot) = \frac{n!}{x_1! \cdots x_k!} \mathbb{1}_{supp}(\underline{x})$

$T(\cdot) = \underline{x}$

$\psi(\cdot) = 0$

$$\Longrightarrow \pi(\theta; \mu, \eta_0) \propto e^{\theta^T \mu}$$

3. Derive the conjugate prior for $(p_1, .., p_{k-1})$ and compute the posterior given $x_1, .., x_n$. Moreover derive the hyperparameters.

from 2.

$$\pi(\theta|\mu) \propto e^{\theta^T\mu} = p_1^{\mu_1} \cdots p_k^{\mu_k} \Rightarrow \text{Dirichlet on } S_{k-1}$$

$$\Rightarrow (p_1, .., p_{k-1}) \sim Dir(\alpha_1, .., \alpha_{k-1}, \alpha_k)$$

Posterior:

$$(p_1, .., p_{k-1}) | \underline{x} \sim Dir(\alpha_1 + n_1, .., \alpha_k + n_k) \qquad n_j = \sum_{i=1}^n x_{ij}$$

4. The predictive?

$$p(\underline{x} | x_1, .., x_n) = \int_{S_{k-1}} f(\underline{x}|\underline{p})\, \pi(\underline{p}|x_1, .., x_n)\, dp_1 \cdots dp_{k-1}$$

3. The marginal distribution for the sample $\boldsymbol{z}$ with the prior found on the previous point is

$$\begin{aligned}
m(\boldsymbol{z}; a, b) &= \int_0^{+\infty} \prod_{i=1}^{5} f(z_i; \lambda)\pi(\lambda) d\lambda \\
&= \frac{1}{\prod_{i=1}^{5} z_i} \int_0^{+\infty} \frac{a^b}{\Gamma(a)} \lambda^{a+n-1} e^{-\lambda\left[b+\sum_{i=1}^{n} \log \frac{z_i}{2}\right]} d\lambda \\
&= \frac{1}{\prod_{i=1}^{5} z_i} \frac{b^a \Gamma(a+n)}{\Gamma(a)\left[b+S_z\right]^{(a+n)}}
\end{aligned}$$

where $S_z = \sum_{i=1}^{5} \log \frac{z_i}{2}$.

We have that $\log m(\boldsymbol{z}; a, b) \propto a \log b - (a+5)\log(b+S_z)$. Then

$$\begin{aligned}
\frac{\partial}{\partial b} \log m(\boldsymbol{z}; a, b) \propto \frac{\partial}{\partial b}\left[a \log b - (a+5)\log(b+S_z)\right] &= \frac{a}{b} - \frac{a+5}{b+S_z} \\
&= \frac{aS_z - 5b}{b(b+S_z)} \geq 0 \Leftrightarrow 5b - aS_z \leq 0 \Leftrightarrow b \leq \frac{a}{5} S_z
\end{aligned}$$

which implies that $b = \frac{a}{5} S_z$ is a point of maximum. With the past observed sample $\boldsymbol{z}$ we have $a = 5$ and $b = S_z = 4.531$.

4. Given the sample at point 3., and the prior specification at point 2., we have

$$S_x = \sum_{i=1}^{n} \log\left(\frac{x_i}{2}\right) = 110.355 - 84\log(2) = 52.131$$

and

$$S_x^* = \sum_{i=1}^{m} \log\left(\frac{x_i^*}{2}\right) = 20.371 - 16\log(2) = 9.281$$

We also implicitly assume that each x_i and each x_j^* are larger or equal than 2. Then the posterior distribution become a $gamma(a_n := a+n, b_n := b + S_z + S_x^*)$ where $a_n := a + n = 5 + 84 = 89$ and $b_n := b + S_x + S_x^* = 4.531 + 52.131 + 9.281 = 65.943$.

Under the quadratic loss function we have that the posterior estimate corresponds to the posterior mean

$$\mathrm{E}[\lambda \mid \boldsymbol{x}, \boldsymbol{x}^*] = \frac{a+n}{b + S_x + S_x^*} = \frac{89}{65.943} \simeq 1.350.$$

5. The Jeffreys' prior $\pi_J(\lambda)$ is proportional to $\sqrt{|I(\lambda)|}$, where $I(\lambda)$ denotes the Fisher's information for a non-censored observation. We have

$$\begin{aligned}
I(\lambda) = \mathrm{E}\left[-\frac{d^2}{d\lambda^2}\ell(\lambda; x)\right] &= \mathrm{E}\left[-\frac{d^2}{d\lambda^2}\left\{\log\lambda - \lambda(b+x)\right\}\right] \\
&= \mathrm{E}\left[\frac{1}{\lambda^2}\right] = \frac{1}{\lambda^2}
\end{aligned}$$

then

$$\pi_J(\lambda) \propto \sqrt{I(\lambda)} \propto \frac{1}{\lambda} I_{(0,+\infty)}(\lambda)$$

and the posterior distribution is proportional to

$$\pi_J(\lambda \mid \boldsymbol{x}, \boldsymbol{x}^*) \propto \lambda^{n-1} e^{-\lambda[S_x + S_x^*]} I_{(0,+\infty)}(\lambda)$$

which is the kernel of a $gamma(n = 84, S_x + S_x^* = 61.412)$ distribution.

6. We want to test

$$H_0 : \lambda \le 1.6 \text{ vs } H_1 : \lambda > 1.6$$

with the prior specified at point 2. The corresponding Bayes factor is

$$BF_{01}(\boldsymbol{x}, \boldsymbol{\delta}) = \frac{\textit{posterior odds}}{\textit{prior odds}} = \frac{\mathbb{P}[\lambda \le 1.6 \mid \boldsymbol{x}, \boldsymbol{x}^*]/\mathbb{P}[\lambda > 1.6 \mid \boldsymbol{x}, \boldsymbol{x}^*]}{\mathbb{P}[\lambda \le 1.6]/\mathbb{P}[\lambda > 1.6]}.$$

A priori we have $\lambda \sim \textit{gamma}(5, 4.531)$, and hence $(2 \times 4.531)\lambda \sim \chi^2_{10}$, so that

$$\mathbb{P}[\lambda \le 1.6] = \mathbb{P}[2 \times 4.531\lambda \le 1.6 \times 2 \times 4.531] = \mathbb{P}\left[\chi^2_{10} \le 14.5\right] = 0.849.$$

Consequently, the prior odds become $\mathbb{P}[\lambda \le 1.6]/\mathbb{P}[\lambda > 1.6] = 5.623$.

For the posterior odds, we use the CLT to approximate the df of the gamma distribution (with a large) by the Gaussian d.f. with same mean and variance. We have that

$$\mathrm{E}[\lambda \mid \boldsymbol{x}, \boldsymbol{x}^*] = \frac{a+n}{b + S_z + S_x^*} = 1.350, \qquad \mathrm{Var}[\lambda \mid \boldsymbol{x}, \boldsymbol{x}^*] = \frac{a+n}{(b + S_z + S_x^*)^2} = 0.020$$

so that $\lambda \sim \textit{gamma}(a+n, b + S_z + S_x^*) \simeq \mathcal{N}(1.340, 0.020)$, and the posterior odds is equal to

$$\mathbb{P}[\lambda \le 1.6 \mid \boldsymbol{x}, \boldsymbol{x}^*]/\mathbb{P}[\lambda > 1.6 \mid \boldsymbol{x}, \boldsymbol{x}^*] \simeq \frac{\Phi\left(\frac{1.6-1.350}{\sqrt{0.020}}\right)}{1 - \Phi\left(\frac{1.6-1.350}{\sqrt{0.020}}\right)} \simeq \frac{\Phi(1.768)}{1 - \Phi(1.768)} \simeq 24.641.$$

The Bayes factor is

$$BF_{01} = \frac{\textit{posterior odds}}{\textit{prior odds}} = \frac{24.641}{5.623} \simeq 4.382$$

and $2 \times \log BF_{01} \simeq 2.955$. There is a evidence in favour of H_0.

7. The posterior predictive survival function is given by

$$\begin{aligned}
1 - F_{X_{5+1}|\boldsymbol{X},\boldsymbol{X}^*}(5; \boldsymbol{x}, \boldsymbol{x}^*) &= \int_0^{+\infty} S_{X_{5+1}}(5; \lambda)\pi(\lambda \mid \boldsymbol{x}, \boldsymbol{x}^*) d\lambda \\
&= \int_0^{+\infty} \frac{2^\lambda}{5^\lambda} \frac{(b + S_x + S_x^*)^{a+n}}{\Gamma(a+n)} \lambda^{a+n-1} e^{-\lambda[b + S_x + S_x^*]} d\lambda = \frac{(b + S_x + S_x^*)^{a+n}}{(b + S_x + S_x^* + \log(5/2))^{a+n}} \\
&= \left(\frac{b_n}{b_n + \log(5/2)}\right)^{a_n}
\end{aligned}$$

and by substituting the opportune quantities, we get

$$1 - F_{X_{5+1}|\boldsymbol{X},\boldsymbol{X}^*}(5; \boldsymbol{x}, \boldsymbol{x}^*) = e^{log(num) - log(den)} \simeq 0.293.$$

∎

Solution of **Exercize 2**.

1. The LHS of (3) can be written as

$$\prod_{j=1}^{k}(1-V_j) = (1-V_k)\prod_{j=1}^{k-1}(1-V_j) = \prod_{j=1}^{k-1}(1-V_j) - V_k\prod_{j=1}^{k-1}(1-V_j) = \prod_{j=1}^{k-1}(1-V_j) - P_k, \quad \text{for } k = 2, \ldots, N-1$$

and this last expression is the RHS of (3). Moreover, equation (3) is also true for $k = N$, since it is equivalent to

$$0 = \prod_{j=1}^{N}(1 - V_j) = \prod_{j=1}^{N-1}(1 - V_j) - P_N \quad (V_N = 1 \text{ a.s.}),$$

that implies it is equivalent to

$$P_N = \prod_{j=1}^{N-1}(1 - V_j) = V_N \prod_{j=1}^{N-1}(1 - V_j)$$

and this is true by the definition of P_N in (2).

Then, by exploiting recursively equation (3) in the previous product, we have

$$P_N = (1 - V_1) - P_{N-1} - P_{N-2} - \cdots - P_2 = 1 - P_{N-1} - P_{N-2} - \cdots - P_2 - P_1,$$

and consequentially $\sum_{j=1}^{N} P_j = 1$ a.s., the statement we want to prove.

Alternatively, one can show that from equation (3):

$$\begin{aligned}
\sum_{k=2}^{N} P_k &= \sum_{k=2}^{N}\left\{\prod_{j=1}^{k-1}(1 - V_j) - \prod_{j=1}^{k}(1 - V_j)\right\} \\
&= (1 - V_1) - \prod_{j=1}^{2}(1 - V_j) + \prod_{j=1}^{2}(1 - V_j) - \prod_{j=1}^{3}(1 - V_j) + \\
&+ \prod_{j=1}^{3}(1 - V_j) - \prod_{j=1}^{4}(1 - V_j) + \cdots + \prod_{j=1}^{N-1}(1 - V_j) - \prod_{j=1}^{N}(1 - V_j) \\
&= (1 - V_1) - \prod_{j=1}^{N}(1 - V_j) = 1 - V_1 \text{ a.s.},
\end{aligned}$$

thanks to $V_N = 1$ a.s.

2. First, note that $\mathrm{E}(P_1) = \mathrm{E}(V_1) = \frac{1}{\alpha+1}$. Since $V_1, \ldots, V_{N-1}$ is an i.i.d. sequence of random variables, for $k = 2, \ldots, N-1$ we have:

$$\mathrm{E}(P_k) = \mathrm{E}\left(V_k \prod_{j=1}^{k-1}(1 - V_j)\right) = \mathrm{E}(V_k) \prod_{j=1}^{k-1}(1 - \mathrm{E}(V_j)) = \frac{1}{\alpha + 1}\frac{\alpha^{k-1}}{(1 + \alpha)^{k-1}}.$$

In force of that, for $k = 1, \ldots, N-2$:

$$\mathrm{E}(P_{k+1}) = \frac{\alpha}{1 + \alpha}\,\mathrm{E}(P_k)\ .$$

Observe that $0 < \eta(\alpha) = \frac{\alpha}{1+\alpha} < 1$, and that it is an increasing function in α. It converges to 0 when α goes to 0, it converges to 1 when α goes to $+\infty$, and it is equal to 0.5 for $\alpha = 1$ (see Figure 1).

Then $\{\mathrm{E}(P_1), \ldots, \mathrm{E}(P_{N-1})\}$ is a decreasing sequence, where the parameter α is tuning its speed. For small values of α, the sequence has a rapidly decreasing behavior, and at the limiting case of $\alpha = 0$ we have $\mathrm{E}(P_k) = 0$ for $k = 2, \ldots, N-1$. When α is large instead, the sequence decreases slowly, and at the limiting scenario of $\alpha = 1$ we have $P_{k+1} = P_k$, $k = 1, \ldots, N-2$. As a final comment, note thatm when $\alpha = 1$, the process described in (2) produces a sequence of "pieces" $P_1, \ldots, P_{N-1}$, such that, on average, at step k the length of P_k is half of the length of P_{k-1}.

∎

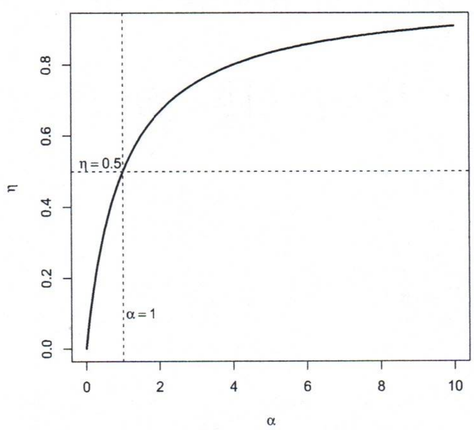


Figura 1: Grafico della funzione $\eta(\alpha)$.

BAYESIAN STATISTICS A. Guglielmi 14.01.2020

Properly justify all your answers. Use the indicator function to denote the support of a distr..

X **Exercise 1** We consider a M/M/1 queue model, describing the queue length of a system with a single server, inter-arrival (or waiting) times $(X_1, \ldots, X_n)$ i.i.d. exponentially distributed with mean $1/\lambda$ and service times $(Y_1, \ldots, Y_n)$ i.i.d. exponentially distributed with mean $1/\mu$. We assume $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ independent. Data consist of $(X_1, Y_1), \ldots, (X_n, Y_n)$, where, for each i, X_i is the waiting time between the arrivals of the $(i-1)$-th and i-th served subjects, and Y_i is the duration of the i-th service.

1. Derive the likelihood with observed data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ (all positive). Then derive a conjugate prior $\pi(\lambda, \mu)$ to this likelihood and find the hyperparameters of the posterior distribution.

2. A priori belief is expressed in terms of a previous observed sample of size $m = 8$, with $\boldsymbol{x}^{old} = (2,\ 1.8,\ 1.7,\ 2.5,\ 2.1,\ 2.0,\ 1.9,\ 2.2)$ and $\boldsymbol{y}^{old} = (1.7,\ 1.5,\ 0.9,\ 1.4,\ 1.8,\ 1.6,\ 1.0,\ 0.8)$. Use the equivalent sample principle to specify the hyperparameters of the conjugate prior π.

3. With hyperparameters at point 2. and actual observed data such that $\sum_{i=1}^n x_i = 22.6$, $\sum_{i=1}^n y_i = 18.4$ and $n = 12$, find the posterior distribution corresponding to the prior π. Provide the posterior estimates of λ and μ under the quadratic loss function.

4. Test the hypotheses $H_0 : \{\lambda = 0.5, \mu = 1\}$ versus $H_1 : \{\lambda \neq 0.5, \mu \neq 1\}$, when the prior density under H_1 is the conjugate prior found at points 1. and 2., by computing the Bayes factor. Write down your conclusion.

When studying a M/M/1 queue system, a key variable is the *stability ratio* λ/μ. In fact, the system is stable if, and only if, λ/μ is strictly smaller than 1.

5. Compute the expectation of the stability ratio and the probability that the system is stable, under the posterior found at point 3.

   (**Hint**: *remember (i) the closeness property of the gamma distributions family wrt the multiplication by a positive constant, that (ii) gamma$(a/2, 1/2) = \chi^2_a$, and that (iii) the ratio of two independent centered chi-square r.v.s $(\chi^2_a/a)/(\chi^2_b/b) \sim F(a, b)$ has the F-distribution with parameters (a, b); the c.d.f. of the $F(40, 40)$ r.v. at $4/3$ is equal to 0.8166.*)

Exercise 2 The Metropolis-Hastings algorithm to sample from a target density $\pi(x)$ (wrt the Lebesgue measure), $x \in E$ is build considering:

- the proposal density $y \mapsto q(x, y)$ on E, for any $x \in E$
- the acceptance probability

$$\alpha(x, y) := \begin{cases} \min\left(\dfrac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right) & \text{if } \pi(x)q(x, y) > 0 \\ 1 & \text{if } \pi(x)q(x, y) = 0 \end{cases}$$

The transition step from time n (the state of the chain being x) to time $n+1$ (the state of the chain being y) is the following: generate a candidate point $Y \sim q(x, y)$, then accept the candidate point with probability $\alpha(x, y)$ as the new state of the chain or remain in the previous state x with probability $1 - \alpha(x, y)$.

If

$$p(x,y) := \begin{cases} q(x,y)\alpha(x,y) & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases},$$

then the transition probability that follows is $P(x,dy) = p(x,y)dy + r(x)\delta_x(dy)$, i.e.

$$P(x,A) = \int_A p(x,y)dy + r(x)I_A(x) = \int_A q(x,y)\alpha(x,y)dy + r(x)I_A(x), \quad A \text{ measurable subset in } E$$

where $r(x) = 1 - \int_E q(x,y)\alpha(x,y)dy$ is the probability that the chain remains at x.

Prove that this chain is reversible wrt π and hence that π is the invariant distribution of this chain.

Exercise 3 Let $(X_1, \ldots, X_n)$ be an iid sample from the Gaussian distribution with mean μ and variance σ^2, with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Let $\sigma := \sqrt{\sigma^2} > 0$, and assume that, a priori, μ and σ are independent,

$$\mu \sim \mathcal{N}(\eta_0, \tau_0^2), \quad \sigma \sim \mathcal{U}((0, \sigma_0)).$$

Derive first the prior induced on σ^2. Then write the full-conditionals of the Gibbs sampler to simulate from the posterior distribution of (μ, σ^2), given $(X_1, \ldots, X_n)$.

Solution of Ex 1.

1. With data $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, the likelihood is:

$$\begin{aligned} L(\lambda, \mu; \boldsymbol{x}, \boldsymbol{y}) &= \prod_{i=1}^{n} \left(\lambda \mu e^{-\lambda x_i - \mu y_i} \right) \mathbf{1}_{(0,+\infty)}(x_i) \mathbf{1}_{(0,+\infty)}(y_i) \\ &= \lambda^n \mu^n e^{-\lambda S_x} e^{-\mu S_y} \mathbf{1}_{(0,+\infty)^n}(\boldsymbol{x}) \mathbf{1}_{(0,+\infty)^n}(\boldsymbol{y}) \end{aligned}$$

   where $S_x = \sum_{i=1}^n x_i$, $S_y = \sum_{i=1}^n y_i$, $\lambda > 0$ and $\mu > 0$. From the previous expression of $L(\lambda, \mu; \boldsymbol{x}, \boldsymbol{y})$ we have that the conjugate prior is the product of two independent gamma distributions. Then, by setting $\pi(\lambda, \mu) = \pi_\lambda(\lambda)\pi_\mu(\mu)$ where

$$\begin{aligned} \pi_\lambda(\lambda \mid a_\lambda, b_\lambda) &= \frac{b_\lambda^{a_\lambda}}{\Gamma(a_\lambda)} \lambda^{a_\lambda - 1} e^{-b_\lambda \lambda} \mathbf{1}_{(0,+\infty)}(\lambda) \\ \pi(\mu \mid a_\mu, b_\mu) &= \frac{b_\mu^{a_\mu}}{\Gamma(a_\mu)} \lambda^{a_\mu - 1} e^{-b_\mu \lambda} \mathbf{1}_{(0,+\infty)}(\mu), \end{aligned}$$

   a posteriori we have that $\pi(\lambda, \mu | data) = \pi_\lambda(\lambda | \boldsymbol{x}) \times \pi_\mu(\mu | \boldsymbol{y})$ and

$$\begin{aligned} \pi(\lambda \mid \boldsymbol{x}) &\propto \lambda^{a_\lambda + n - 1} e^{-\lambda(b_\lambda + S_x)} \mathbf{1}_{(0,+\infty)}(\lambda) \\ \pi(\mu \mid \boldsymbol{y}) &\propto \mu^{a_\mu + n - 1} e^{-\mu(b_\mu + S_y)} \mathbf{1}_{(0,+\infty)}(\mu), \end{aligned}$$

   where the previous expressions coincide with the kernels of two gamma distributions, respectively $\lambda \mid \boldsymbol{x} \sim gamma(a_\lambda + n, b_\lambda + S_x) = gamma(\widetilde{a}_\lambda, \widetilde{b}_\lambda)$ and $\mu \mid \boldsymbol{y} \sim gamma(a_\mu + n, b_\mu + S_y) = gamma(\widetilde{a}_\mu, \widetilde{b}_\mu)$.

2. The maximum likelihood estimators for λ and μ, which are independently distributed, correspond to

$$\hat{\lambda}_{ML} = \arg\max_\lambda \{\ell(\lambda, \mu; \boldsymbol{x}, \boldsymbol{y})\} = \left\{ \lambda : \frac{n}{\lambda} - S_x = 0 \right\} = \frac{S_x}{n}$$

(maximum point, since the second derivative is negative in the point), and similarly we have $\hat{\mu}_{ML} = \frac{S_y}{n}$. The posterior mean of λ is

$$\mathrm{E}[\lambda \mid \boldsymbol{x}] = \frac{b_\lambda + S_x}{a_\lambda + n} = \frac{a_\lambda}{a_\lambda + n}\frac{b_\lambda}{a_\lambda} + \frac{n}{a_\lambda + n}\frac{S_x}{n}.$$

By the equivalent sample principle, we interpret the parameter a_λ as the sample size of an equivalent sample, and b_λ as the sum of the observations in the equivalent sample. In this case, we have $a_\lambda = 8$ and $b_\lambda = 16.2$. Similarly we find $a_\mu = 8$ and $b_\mu = 10.7$.

3. As shown in point 1., the posterior distribution is the product of two gamma densities. By specifying the hyperparameters according to the equivalent sample principle, and conditionally to the observed sample, we have

$$\lambda \mid \boldsymbol{x} \sim gamma(a_\lambda + n, b_\lambda + S_x) \stackrel{d}{=} gamma(20, 38.8)$$
$$\mu \mid \boldsymbol{y} \sim gamma(a_\mu + n, b_\mu + S_y) \stackrel{d}{=} gamma(20, 29.1)$$

The point estimate minimizing the quadratic loss function a posteriori is the the posterior mean, that yields

$$\mathrm{E}(\lambda|data) = \frac{a_\lambda + n}{b_\lambda + S_x} = \frac{20}{38.8} \simeq 0.5155$$
$$\mathrm{E}(\mu|data) = \frac{a_\mu + n}{b_\mu + S_x} = \frac{20}{29.1} \simeq 0.6873.$$

4. We want to test the hypotheses $H_0 : \lambda = 0.5, \mu = 1$ vs $H_1 : \lambda \neq 0.5, \mu \neq 1$, a point null H_0 against a "diffuse" H_1. The corresponding Bayes factor is

$$BF_{01}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\prod_{i=1}^n f(x_i, y_i; \lambda = 0.5, \mu = 1)}{m_1(\boldsymbol{x}, \boldsymbol{y})},$$

where

$$\begin{aligned} m_1(\boldsymbol{x}, \boldsymbol{y}) &= \int_{(0,+\infty)} \int_{(0,+\infty)} \prod_1^n f(x_i; \theta)\pi(\lambda, \mu) d\lambda d\mu \\ &= \frac{b_\lambda^{a_\lambda} b_\mu^{a_\mu}}{\Gamma(a_\lambda)\Gamma(a_\mu)} \int_0^{+\infty} \int_0^{+\infty} \lambda^{a_\lambda + n - 1} \mu^{a_\mu + n - 1} \mathrm{e}^{-\lambda(b_\lambda + S_x) - \mu(b_\mu + S_y)} d\lambda d\mu \\ &= \frac{b_\lambda^{a_\lambda}}{\Gamma(a_\lambda)} \frac{\Gamma(a_\lambda + n)}{(b_\lambda + S_x)^{a_\lambda + n}} \times \frac{b_\mu^{a_\mu}}{\Gamma(a_\mu)} \frac{\Gamma(a_\mu + n)}{(b_\mu + S_y)^{a_\mu + n}} \simeq 1.914916\ 10^{-9} \times 2.18713\ 10^{-8} \simeq 4.188169 \times 10^{-17}, \end{aligned}$$

while the numerator of the BF is equal to $\prod_{i=1}^n f(x_i, y_i; \lambda = 0.5, \mu = 1) = (0.5)^{12} \mathrm{e}^{(-0.5 S_x - S_y)} \simeq 3.083855 \times 10^{-17}$, so that

$$\begin{aligned} BF_{01}(\boldsymbol{x}, \boldsymbol{y}) &= \frac{\Gamma(a_\lambda)\Gamma(a_\mu)(b_\lambda + S_x)^{a_\lambda + n}(b_\mu + S_y)^{a_\mu + n} 0.5^n 1^n \mathrm{e}^{-0.5 S_x - 1 S_y}}{b_\lambda^{a_\lambda} b_\mu^{a_\mu} \Gamma(a_\lambda + n)\Gamma(a_\mu + n)} \\ &= \frac{3.083855 \times 10^{-17}}{4.188169 \times 10^{-17}} \simeq 0.7363 \Rightarrow 2 \log BF_{01}(\boldsymbol{x}, \boldsymbol{y}) \simeq -0.6122. \end{aligned}$$

We conclude that there is weak evidence in favour of H_1.

5. The posterior expectation of the stability ratio is

$$\mathrm{E}\left(\frac{\lambda}{\mu} \mid (x_1, y_1), \ldots, (x_n, y_n)\right) = \mathrm{E}(\lambda \mid \boldsymbol{x})\, \mathrm{E}\left(\frac{1}{\mu} \mid \boldsymbol{y}\right)$$
$$= \frac{a_\lambda + n}{b_\lambda + S_x} \frac{b_\mu + S_y}{a_\mu + n - 1} = \frac{20}{38.8} \frac{29.1}{19} = \frac{15}{19} \simeq 0.7895.$$

Remember that, if $\lambda \sim gamma(a, b)$, then $Y = c\lambda \sim gamma(a, \frac{b}{c})$ with a positive constant c; in fact

$$f_Y(y) = f_\lambda(\frac{y}{c})\frac{1}{c} = \frac{1}{c}\frac{b^a}{\Gamma(a)}\left(\frac{y}{c}\right)^{a-1} \mathrm{e}^{-b\frac{y}{c}} \mathbf{1}_{(0,+\infty)}(\frac{y}{c}) \propto y^{a-1} \mathrm{e}^{\frac{b}{c}y} \mathbf{1}_{(0,+\infty)}(y).$$

Hence, if $c = 2b$, then

$$2b\lambda \sim gamma(a, \frac{1}{2}) = gamma(\frac{2a}{2}, \frac{1}{2}) = \chi^2_{2a}.$$

Moreover, since a posteriori $\lambda \sim gamma(\widetilde{a}_\lambda, \widetilde{b}_\lambda)$ and $\mu \sim gamma(\widetilde{a}_\mu, \widetilde{b}_\mu)$, with λ and μ independent, so that $2\widetilde{b}_\lambda \lambda \sim \chi^2_{2\widetilde{a}_\lambda}$ and $2\widetilde{b}_\mu \mu \sim \chi^2_{2\widetilde{a}_\mu}$, then we have

$$\frac{2\widetilde{b}_\lambda \frac{\lambda}{2\widetilde{a}_\lambda}}{2\widetilde{b}_\mu \frac{\mu}{2\widetilde{a}_\mu}} = \frac{(b_\lambda + S_x)(a_\mu + n)\lambda}{(b_\mu + S_y)(a_\lambda + n)\mu} = \frac{(b_\lambda + S_x)\lambda}{(b_\mu + S_y)\mu} \sim F(2\widetilde{a}_\lambda, 2\widetilde{a}_\mu) = F(40, 40),$$

since $\widetilde{a}_\lambda = a_\lambda + n = \widetilde{a}_\mu = a_\mu + n = 20$.

The posterior probability that the system is stable is:

$$P\left(\frac{\lambda}{\mu} < 1 \mid (X_1, Y_1), \ldots, (X_n, Y_n)\right)$$
$$= P\left(\frac{(b_\lambda + S_x)(a_\mu + n)\lambda}{(b_\mu + S_y)(a_\lambda + n)\mu} < \frac{2(b_\lambda + S_x)(a_\mu + n)}{2(b_\mu + S_y)(a_\lambda + n)} \mid (X_1, Y_1), \ldots, (X_n, Y_n)\right)$$
$$= P\left(F(2(a_\mu + n), 2(a_\lambda + n)) < \frac{b_\lambda + S_x}{b_\mu + S_y}\right)$$
$$= P\left(F(40, 40) < \frac{38.8}{29.1}\right) = P\left(F(40, 40) < \frac{4}{3}\right) = 0.8166.$$

Here $F(2(a_\mu + n), 2(a_\lambda + n)) = F(40, 40)$ denotes the random variable with F- distribution with parameters $(40, 40)$.

∎

Solution of Ex 2.
We first prove the reversibility condition, i.e. that

$$p(x, y)\pi(x) = p(y, x)\pi(y) \Leftrightarrow q(x, y)\alpha(x, y)\pi(x) = q(y, x)\alpha(y, x)\pi(y).$$

In fact, if $\dfrac{\pi(y)q(y, x)}{\pi(x)q(x, y)} < 1$, then the LHS above is equal to .... On the other hand, if $\dfrac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \geq 1$, then the LHS above is equal to ....

Now, let us prove that π is invariant for the MH chain: For any measurable subset A of E we have:

$$\int_E P(x, A)\pi(x)dx = \int_E \left(\int_A p(x, y)dy + r(x)I_A(x)\right)\pi(x)dx = \int_A dy \int_E p(x, y)\pi(x)dx + \int_A r(x)\pi(x)dx$$

Thanks to the reversibility condition, the right hand-side of the formula above is equal to

$$\int_A dy \int_E p(y,x)\pi(y)dx + \int_A r(x)\pi(x)dx = \int_A \pi(y)(1-r(y))dy + \int_A r(y)\pi(y)dy = \int_A \pi(y)dy = \pi(A).$$

This means that π is invariant for the MH chain. ∎

Solution of Ex 3.
First of all, if $Y := \sigma^2$, then, for $y > 0$,

$$\pi_Y(y) = \pi_\sigma(\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}\frac{1}{\sigma_0}\mathbf{1}_{(0,\sigma_0)}(\sqrt{y}) = \frac{1}{2\sigma_0\sqrt{y}}\mathbf{1}_{(0,\sigma_0^2)}(y).$$

The joint distribution of data and parameters, in this case, is as follows:

$$\mathcal{L}(\boldsymbol{X},\mu,\sigma^2) = \mathcal{L}(\boldsymbol{X}|\mu,\sigma^2)\pi_\mu(\mu)\pi_{\sigma^2}(\sigma^2) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sigma^2}\right)^{n/2} \mathrm{e}^{-(\sum_1^n(x_i-\mu)^2)/(2\sigma^2)}\pi_\mu(\mu)\pi_{\sigma^2}(\sigma^2).$$

The full-conditional of σ^2 is given by:

$$[\sigma^2|\mu,\boldsymbol{x}] \propto (\sigma^2)^{-\frac{1}{2}-\frac{n}{2}}\mathrm{e}^{-(\sum_1^n(x_i-\mu)^2)/(2\sigma^2)}\mathbf{1}_{(0,\sigma_0^2)}(\sigma^2) = \left(\frac{1}{\sigma^2}\right)^{\frac{n+1}{2}-1+1}\mathrm{e}^{-\frac{\sum_1^n(x_i-\mu)^2}{2\sigma^2}}\mathbf{1}_{(0,\sigma_0^2)}(\sigma^2)$$
$$\sim Trunc - invgamma\left(\frac{n-1}{2}, \frac{\sum_1^n(x_i-\mu)^2}{2}\right)$$

that is a inv-gamma distribution with parameters $\frac{n-1}{2}$, $\frac{\sum_1^n(x_i-\mu)^2}{2}$ truncated on the interval $(0,\sigma_0^2)$.
The full-conditional of μ is given by:

$$[\mu|\sigma^2,\boldsymbol{x}] \propto \mathrm{e}^{-\frac{(\mu-\eta_0)^2}{2\tau_0^2}}\mathrm{e}^{-(\sum_1^n(x_i-\mu)^2)/(2\sigma^2)} \propto \exp\{-\frac{1}{2}\frac{\sigma^2+n\tau_0^2}{\tau_0^2\sigma^2}(\mu^2-2\mu\frac{\sigma^2\eta_0+n\tau_0^2\bar{x}}{\sigma^2+n\tau_0^2})\}$$
$$= \mathcal{N}\left(\frac{\sigma^2\eta_0+n\tau_0^2\bar{x}}{\sigma^2+n\tau_0^2}, \frac{\sigma^2\tau_0^2}{\sigma^2+n\tau_0^2}\right).$$

∎

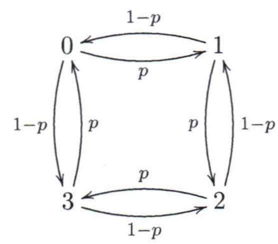
BAYESIAN STATISTICS   A. Guglielmi   13.06.2019

Properly justify all your answers. Use the indicator function to denote the support of a distr..

Exercise 1 A particle moves among four states (positions) $\mathcal{S} = \{0,1,2,3\}$. Let $\{X_n, n \geq 0\}$ be the sequence of the states of the particle, X_n being the state of the particle at time n, and let $p \in (0,1)$ be an unknown parameter. The particle moves according to the following rules:

- If the particle state at time n is $x \in \{0,1,2\}$, then at time $n+1$ the particle will be in state $x+1$ with probability p.
- If the particle state at time n is $x \in \{1,2,3\}$, then at time $n+1$ the particle will be in state $x-1$ with probability $1-p$.
- If the particle is at state 3 at time n, at time $n+1$ the particle will be in state 0 with probability p.
- If the particle is at state 0 at time n, at time $n+1$ the particle will be in state 3 with probability $1-p$.

Here is is a diagram describing the transition rules:



We observe the particle at times $0, 1, \ldots, n$, and let $x_0, x_1, \ldots, x_n$ be the observed sequence of states. Moreover, denote by n^+ the observed total number of clockwise (see the diagram) transitions, i.e. the number of transitions from 0 to 1, from 1 to 2, from 2 to 3, and from 3 to 0, and by n^- the observed total number of counterclockwise transitions, i.e. the number of transitions from 1 to 0, from 2 to 1, from 3 to 2 and from 0 to 3.

The aim is to make inference on the unknown parameter p, given this observed sequence of particle positions: LO ZERO NON RIENTRA IN $\underline{x}_n = (x_1, .., x_n)$

$$x_0 = 0, 1, 0, 3, 2, 3, 2, 1, 2, 3, 2, 1, 2, 3, 2, 1.$$ QUI $n = 15$

1. Conditioning to $X_0 = x_0 = 0$, write the likelihood of data as a function of n^+ and n^-, i.e. compute
$$L_{x_0}(p; x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | X_0 = x_0, p)$$
(**Hint**: *factorize* $\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | X_0 = x_0, p)$ *using the product rule*)

2. Derive the conjugate prior to this likelihood, and find the hyperparameters of the posterior. Note that the prior depends on positive hyperparameters α and β.

3. From a past experiment we have observed the particle for $m = 10$ transitions with $m^+ = 3$ clockwise transitions. Use the equivalent sample approach to fix the parameters of the conjugate prior π_1 found at point 2. and denote by π_1 the conjugate prior with hyperparameters here derived.

4. Under prior π_1, compute posterior mean and variance of p with the available data.

5. Assume now that, instead of information at point 3., we only know that we have observed 40% of transitions from 0 to 1 in past experiments. Within the conjugate class of priors, find 3 couples of hyperparameters α, β such that the prior predictive probability to have a transition from 0 to 1 (i.e. $\mathbb{P}(X_1 - X_0 = 1 | X_0 = 0)$) is equal to the observed percentage in past experiments, when $\alpha \in \{0.5, 1, 2\}$. Are posterior mean and variance of p robust under these three choices of hyperparameters?

= media e varianze cambiano poco al variare di questi parametri? (NB. considere TUTTE le coppie)

6. Consider now the four models

$$M_j = \begin{cases} \text{likelihood computed at point 1.} \\ \pi(p|M_j), \quad p \in (0,1) \end{cases} \qquad j \in \{1,2,3,4\}$$

where $\pi(p|M_1) = \pi_1$ while $\pi(p|M_2)$, $\pi(p|M_3)$, and $\pi(p|M_4)$ are the priors computed at point 5. If $\mathbb{P}(M = M_j) = 1/4$ for all $j \in \{1, \ldots, 4\}$, compute the posterior probability that $\mathbb{P}(M = M_j|\text{data})$. Which model do you choose and why?

Remember that the beta special function is defined as

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx, \qquad \alpha > 0, \beta > 0\ . \tag{1}$$

The following table could be useful in your calculations:

(α, β)	(0.5,0.75)	(1,1.5)
log B(α, β)	0.874	-0.405
log B$(\alpha+6, \beta+9)$	-10.682	-11.561

Exercise 2 Compute the marginal distribution of a sample $(X_1, X_2, \ldots, X_n)$ from a Dirichlet process P with parameters (α, P_0), where $\alpha > 0$ is the total mass and P_0 is centering probability measure. Provide the proof.

Solution of Ex 1.

1. The sequence $(X_n)_{n\geq 0}$ is a Markov chain; hence:

$$\begin{aligned} L_{x_0}(p; x_1, \ldots, x_n) &= \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | X_0 = x_0, p) \\ &= \mathbb{P}(X_1 = x_1 | X_0 = x_0, p)\mathbb{P}(X_2 = x_2 | X_1 = x_1, p) \cdots \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, p). \end{aligned} \tag{2}$$

For $j = 1, \ldots, n$, each factor in (2) is given by p if the transition from x_{j-1} to x_j is clockwise, or $1 - p$ if the transition is counterclockwise. We have:

$$p(x_1, \ldots, x_n | x_0, p) = p^{n^+}(1-p)^{n^-} = p^{n^+}(1-p)^{n-n^+}\mathbf{1}_{(0,1)}(p).$$

2. The likelihood computed at point 1. is analytically equal to the Bernoulli likelihood (with success probability p). The conjugate prior for such model is the beta(α, β) distribution, with $\alpha, \beta > 0$. If $\pi(p)$ denotes its density, then:

$$\pi(p|x_0, \ldots, x_n) \propto p(x_1, \ldots, x_n | x_0, p)\pi(p) = p^{n^+}(1-p)^{n^-} p^{\alpha-1}p^{\beta-1}\mathbf{1}_{(0,1)}(p) = p^{n^+ + \alpha - 1}(1-p)^{n^- + \beta - 1}\mathbf{1}_{(0,1)}(p).$$

It is clear that the posterior density of p is beta$(n^+ + \alpha, n^- + \beta)$.

3. We have:

$$\mathrm{E}(p|x_1,\dots,x_n) = \frac{n^+ + \alpha}{n+\alpha+\beta} = \frac{n}{n+\alpha+\beta}\frac{n^+}{n} + \frac{\alpha+\beta}{n+\alpha+\beta}\frac{\alpha}{\alpha+\beta}.$$

By the equivalent sample approach, parameter $\alpha+\beta$ represents the "prior sample size", while $\alpha/(\alpha+\beta)$ represents the "prior mean". Fixing these quantities equal to 10 and 3/10 respectively, we have $\alpha_1 = 3$ and $\beta_1 = 7$. We denote by π_1 the beta($\alpha_1 = 3, \beta_1 = 7$) density.

4. Data contain $n^+ = 6$ clockwise transitions and $n^- = 9$ counterclockwise, with $n = 15$. Posterior distribution of p, when the prior is π_1, is beta(α_1+6, β_1+9)=beta(9, 16). Remember that, if $p \sim$Beta(α, β), then

$$\mathrm{E}(p) = \frac{\alpha}{\alpha+\beta} \quad \text{e} \quad \mathrm{Var}(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Hence

$$\mathrm{E}_{\pi_1}(p|x_0,\dots,x_n) = \frac{9}{25} \simeq 0.36, \quad \mathrm{Var}(p|x_0,\dots,x_n) = \frac{72}{8125} \simeq 0.0089.$$

5. Observe that

$$\begin{aligned}\mathbb{P}(X_1 - X_0 = 1|X_0 = 0) = \mathbb{P}(X_1 = 1|X_0 = 0) &= \int_0^1 \mathbb{P}(X_1 = 1|p, X_0 = 0)\pi(p|X_0 = 0)dp \\ &= \int_0^1 \mathbb{P}(X_1 = 1|p, X_0 = 0)\pi(p)dp = \int_0^1 p\pi(p)dp = \mathrm{E}_\pi(p) = \frac{\alpha}{\alpha+\beta},\end{aligned}$$

where π is the beta(α, β) density. From $\frac{\alpha}{\alpha+\beta} = 0.4$, we get $\beta = \frac{3}{2}\alpha$, so that when $\alpha_2 = 0.5$ we have $\beta_2 = \frac{3}{4}$, for $\alpha_3 = 1$ we have $\beta_3 = \frac{3}{2}$ and finally for $\alpha_4 = 2$ we have $\beta_4 = 3$. In the following table we report posterior mean and variance for each of the four priors we have considered so far.

Prior hyperpar	Poster Hyperpar	PosterMean	PosterVar
$\alpha_1 = 3, \beta_1 = 7$	$\alpha_1^* = 9, \beta_1^* = 16$	0.36	0.0089
$\alpha_1 = 0.5, \beta_1 = 0.75$	$\alpha_2^* = 6.5, \beta_2^* = 9.75$	0.4	0.0139
$\alpha_1 = 1, \beta_1 = 1.5$	$\alpha_3^* = 7, \beta_3^* = 10.5$	0.4	0.0130
$\alpha_1 = 2, \beta_1 = 2$	$\alpha_4^* = 8, \beta_4^* = 12$	0.4	0.0114

Both posterior mean and variance are robust in this case.

6. Let $\boldsymbol{x} = (x_1,\dots,x_n)$; for all $j \in \{1,\dots,4\}$ the posterior distribution of p is given by

$$\pi(p|\boldsymbol{x}, x_0, M_j) = \frac{p(x_1,\dots,x_n|p, x_0, M_j)\pi(p|M_j)}{m(x_1,\dots,x_n|x_0, M_j)},$$

where $p(x_1,\dots,x_n|x_0, p, M_j)$ is the conditional joint distribution computed at point 1.; this distribution does not change with j. Moreover,

$$\begin{aligned} m(x_1,\dots,x_n|x_0, M_j) &= \int_0^1 p(x_1,\dots,x_n|x_0, p)\pi_j(p)dp = \\ &= \frac{1}{B(\alpha_j, \beta_j)}\int_0^1 p^{n^+ + \alpha_j - 1}(1-p)^{n^- + \beta_j - 1}dp = \frac{B(\alpha_j + n^+, \beta_j + n^-)}{B(\alpha_j, \beta_j)}\end{aligned}$$

where $B(\cdot,\cdot)$ is the beta function in (1). By Bayes' theorem we have that

$$\mathbb{P}(M = M_j|x_0,\dots,x_n) = \frac{m(x_1,\dots,x_n|x_0, M_j)\mathbb{P}(M = M_j)}{m(x_1,\dots,x_n|x_0)}, \quad j = 1, 2, 3, 4\ ,$$

where $m(x_1, \dots, x_n|x_0) = \sum_{j=1}^4 m(x_1, \dots, x_n|x_0, M_j)\mathbb{P}(M = M_j)$. We choose model M_j with maximum posterior probability. Since $\mathbb{P}(M = M_j) = 1/4$ for all j, then

$$\mathbb{P}(M = M_j|x_0, \dots, x_n) \propto m(x_1, \dots, x_n|x_0, M_j)$$

Now, it is straightforward to compute $\log m(x_1, \dots, x_n|x_0, M_j)$ from the values in this table

(α, β)	(3,7)	(0.5,0.75)	(1,1.5)	(2,3)
log $\mathrm{B}(\alpha, \beta)$	-5.529	0.874	-0.405	-2.485
log $\mathrm{B}(\alpha + 6, \beta + 9)$	-16.281	-10.682	-11.561	-13.312

and obtain that

$\log m(\boldsymbol{x}|x_0, M_1)$	-10.752
$\log m(\boldsymbol{x}|x_0, M_2)$	-11.556
$\log m(\boldsymbol{x}|x_0, M_3)$	-11.156
$\log m(\boldsymbol{x}|x_0, M_4)$	-10.827

We choose M_1 among the four models, since it achieves the maximum value of the marginal $m(\boldsymbol{x}|x_0, M_j)$ and consequently of the posterior probability $\mathbb{P}(M = M_j|x_0, \dots, x_n)$.

▮

Solution of Ex 2.
If

$$\begin{aligned} X_1, \dots, X_n|P &\overset{iid}{\sim} P \\ P|\alpha, P_0 &\sim DP(\alpha, P_0), \end{aligned} \tag{3}$$

then the marginal distribution of $(X_1, \dots, X_n)$ can be found as

$$\begin{aligned} \mathcal{L}(X_1, \dots, X_n) &= \mathcal{L}(X_1) \times \mathcal{L}(X_2|X_1) \times \dots \times \mathcal{L}(X_n|X_1, \dots, X_{n-1}) \\ &= \prod_{i=1}^n \left(\frac{\alpha P_0(dX_i) + \sum_{j=1}^{i-1} \delta_{X_j}(dX_i)}{\alpha + i - 1} \right) \end{aligned}$$

with the convention that the element in the brackets is P_0 when $i = 1$. This expression can be obtained by marginalizing out the random measure P from (3).

▮

BAYESIAN STATISTICS A. Guglielmi X 14.02.2019

Properly justify all your answers. Use the indicator function to denote the support of a distr..

X **Exercise 1** Data below represent lifetimes of kitchen mixers (expressed in years) used in a TV food show and then sold on the internet:

4.77 0.29 0.71 2.26 5.02 3.58 4.85 1.39 0.44 2.63 .

To make inference on these failure times we use the following Bayesian model:

$$X_1, \dots, X_n | \theta \overset{i.i.d}{\sim} \text{Exp}(\theta), \text{ with } \theta > 0, \tag{1}$$
$$\theta \sim \pi(\theta),$$

where $\pi(\theta)$ is a prior density (with support $(0, +\infty)$) and θ represents the failure rate of the exponential distribution.

1. Derive Jeffreys prior π_1 and a posterior estimate of θ when the prior for model (1) is π_1.
2. Derive a conjugate prior π_2 for (1); verify that this prior has two parameters, the shape and the rate, such that $\mathrm{E}_{\pi_2}(\theta) = shape/rate$. Find the hyperparameters of the posterior distribution.
3. Derive the Bayesian estimator of θ minimizing the posterior expected value of the quadratic loss function, using the posterior derived at point 2.; interpret shape and rate parameters using the prior sample principle.
4. Derive the (prior) marginal density m_{X_1} of X_1 under π_2 as a prior for θ. In addition, compute its quantile function
$$Q_{X_1}(p) = F_{X_1}^{-1}(p) \quad \text{for } p \in (0, 1),$$
where F_{X_1} denotes the prior marginal distribution function of X_1.
5. To set the hyperparameters in π_2, use the *empirical Bayes* approach as follows: fix the shape parameter equal to 10, and derive the rate parameter such that the difference between $Q_{X_1}(0.975)$ and $Q_{X_1}(0.025)$ is equal to the range of the dataset under investigation. Compute the hyperparameters of the posterior distribution of θ given available data, and the posterior mean of θ.

   (**Hint**: *Check out that you know what is the* ***range of a sample!***)
6. Test the hypotheses that the conditional mean failure time of the kitchen mixer under (1) is larger than 2.5 years (H_0) versus the hypothesis that the conditional mean is less than this bound. Use π_2 as a prior on θ.

   (**Hint**: *Useful to know that the prior df, evaluated at 0.4, is 0.0123. To compute the posterior df, use a proper approximation.*)

In real world problems in Survival Analysis, we typically have right-censored data. A standard dataset is represented as

$$\text{data} = (X_1, \dots, X_m, X_{m+1}^+ \dots, X_{m+k}^+), \tag{2}$$

where X_{m+i}^+, for $i = 1, \dots, k$, denote right-censored observations, when the lifetimes of subjects $m+1, \dots, m+k$ are larger than thresholds $X_{m+1}^+ \dots, X_{m+k}^+$, respectively. For data in (2), assume now model (1) under the conjugate prior π_2 with general shape and rate parameters.

7. Using a *data augmentation* technique, derive the full-conditionals of a Gibbs sampler to simulate from the posterior

$$\mathcal{L}(\theta, Z_1, \ldots, Z_k|\text{data}),$$

where $Z_1, \ldots, Z_k$ represent the *augmented* failure times of the k subjects corresponding to right-censored observations.

(**Hint**: *It is of paramount importance to derive first the conditional law* $\mathcal{L}(Z_j|\theta, X^+_{m+j})$ *for any* $j = 1, \ldots, k$.)

Solution of Ex. 1.

1. For $x_i > 0$, $i = 1, \ldots, n$, the likelihood function of model (1) is

$$L(\theta, \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \theta) = \theta^n \mathrm{e}^{-\theta \sum_1^n x_i} = \exp\left\{-\sum_{i=1}^{n} x_i\theta + n\log(\theta)\right\}, \qquad \theta > 0. \tag{3}$$

The Fisher information has the following expression:

$$I(\theta) = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X, \theta)\right] = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\{-\theta X + \log(\theta)\}\right] = \frac{1}{\theta^2}.$$

Hence, Jeffreys prior is such that

$$\pi_J(\theta) \propto \sqrt{I(\theta)} = \frac{1}{\theta}\mathbf{1}_{(0,\infty)}(\theta),$$

that is an improper prior! The resulting posterior is

$$\pi_J(\theta|\underline{x}) \propto \theta^{n-1}\mathrm{e}^{-\theta\sum_{i=1}^n x_i}\mathbf{1}_{(0,+\infty)}(\theta), \tag{4}$$

which is the $gamma(n, \sum_{i=1}^n x_i)$ distribution, i.e. $gamma(10, 25.94)$ under available data. An estimate for θ is the posterior mean $\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i} = \frac{10}{25.94} \simeq 0.3855$.

2. It is straightforward to see from the expression of the likelihood (3) that a conjugate prior has the be proportional to $\theta^{\alpha-1}\mathrm{e}^{-\beta\theta}\mathbf{1}_{(0,\infty)}(\theta)$, so that the posterior, proportional to the likelihood times the prior, is such that

$$\theta^n \mathrm{e}^{-\theta\sum_1^n x_i}\theta^{\alpha-1}\mathrm{e}^{-\beta\theta}\mathbf{1}_{(0,\infty)}(\theta) = \theta^{\alpha+n}\mathrm{e}^{-\theta(\beta+\sum_1^n x_i)}\mathbf{1}_{(0,\infty)}(\theta).$$

Summing up, the conjugate prior π_2 for θ is the $gamma(\alpha, \beta)$ distribution, $\alpha, \beta > 0$ and the corresponding posterior $\pi_2|\boldsymbol{x}$ is the $gamma(\alpha + n, \beta + \sum_1^n x_i)$ distribution. Since $\mathrm{E}_{\pi_2}(\theta) = \alpha/\beta$, parameter α is the *shape*, while β is the *rate*.

3. The Bayesian estimator of θ minimizing the posterior expected value of the quadratic loss function is the posterior mean of θ, that is

$$\mathrm{E}(\theta|\underline{x}) = \left(\frac{\sum_{i=1}^n x_i + \beta}{n+\alpha}\right)^{-1} = \left(\frac{n}{n+\alpha}\bar{x} + \alpha\frac{\beta/\alpha}{n+\alpha}\right)^{-1} = \left(\frac{n}{n+\alpha}\bar{x} + \mathrm{E}(\theta)\frac{\alpha}{n+\alpha}\right)^{-1}.$$

We interpret α as the prior sample size, while β is the total sum of prior sample data.

4. The marginal (prior) density m_{X_1} of X_1 in (1) under π_2 as a prior for θ has the following expression for $x > 0$:

$$m_{X_1}(x) = \int_0^{+\infty} f(x;\theta)\pi_2(\theta)d\theta = \int_0^{+\infty} \theta e^{-\theta x}\frac{1}{\Gamma(\alpha)}\beta^\alpha\theta^{\alpha-1}e^{-\beta\theta}d\theta = \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+1)}{(x+\beta)^{\alpha+1}} = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}.$$

The distribution function of $m_{X_1}(x)$, when $x > 0$, is:

$$F_{X_1}(x) = \int_0^x m_{X_1}(u)du = \int_0^x \frac{\alpha\beta^\alpha}{(u+\beta)^{\alpha+1}}du = \alpha\beta^\alpha\left[-\frac{1}{\alpha(u+\beta)^\alpha}\right]_0^x = 1 - \frac{\beta^\alpha}{(x+\beta)^\alpha}.$$

Fix now $p \in (0,1)$; the quantile function $Q_{X_1}(p)$ is computed solving equation $F_{X_1}(x) = p$ with respect to x. We have:

$$1 - \frac{\beta^\alpha}{(x+\beta)^\alpha} = p \Rightarrow \frac{\beta}{x+\beta} = (1-p)^{1/\alpha} \Rightarrow Q_{X_1}(p) = \beta\left(\frac{1}{(1-p)^{1/\alpha}} - 1\right).$$

5. Let $p_1 = 0.975$ e $p_2 = 0.025$. Then

$$I := Q_{X_1}(p_1) - Q_{X_1}(p_2) = \beta\left(\frac{1}{(1-p_1)^{1/\alpha}} - \frac{1}{(1-p_2)^{1/\alpha}}\right),$$

while the range of the data is $x_{(n)} - x_{(1)} = 5.02 - 0.29 = 4.73$. Fixing $\alpha = 10$, from equation $I = 4.73$ we get $\beta \simeq 10.663$. Hence, a posteriori $\theta|\boldsymbol{x} \sim gamma(\alpha_n, \beta_n) = gamma(20, 36.603)$. The posterior mean of θ is $\alpha_n/\beta_n = 0.5464$.

6. The condition to be tested as H_0 is whether $\mathrm{E}(X_1|\theta) = 1/\theta$ is larger than 2.5, so that

$$H_0\colon\ \theta < \tfrac{1}{2.5} = 0.4 \text{ vs } H_1\colon\ \theta \geq 0.4,$$

The Bayes factor is the ratio between posterior and prior odds:

$$\mathrm{BF}_{01} = \frac{\frac{\mathbb{P}(H_0|\boldsymbol{x})}{\mathbb{P}(H_1|\boldsymbol{x})}}{\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}} = \frac{\frac{\mathbb{P}(\theta<0.4|\boldsymbol{x})}{\mathbb{P}(\theta\geq0.4|\boldsymbol{x})}}{\frac{\mathbb{P}(\theta<0.4)}{\mathbb{P}(\theta\geq0.4)}} = \frac{\frac{F_{\pi_2}(0.4|\boldsymbol{x})}{1-F_{\pi_2}(0.4|\boldsymbol{x})}}{\frac{F_{\pi_2}(0.4)}{1-F_{\pi_2}(0.4)}}$$

Since $F_{\pi_2}(0.4) = 0.0123$, the prior odds are computed as

$$\frac{P(H_0)}{P(H_1)} = \frac{0.0123}{1-0.0123} = 0.0124164.$$

As far as the posterior of θ is concerned, note that, though $\alpha_n = 20$ is not extremely large, the Gaussian approximation is still good, i.e.

$$\theta|\boldsymbol{x} \sim gamma(\alpha_n, \beta_n) = gamma(20, 36.603) \approx \mathcal{N}(\frac{\alpha_n}{\beta_n}, \frac{\alpha_n}{\beta_n^2}) = \mathcal{N}(0.5464, 0.0149),$$

so that $F_{\pi_2}(0.4|\boldsymbol{x}) \simeq \Phi(\frac{0.4-0.5464}{\sqrt{0.0149}}) \simeq \Phi(-1.1994) = 1 - \Phi(1.1994) \simeq 0.1152$ (the exact value from the gamma distribution is 0.1057), and the posterior odds are

$$\frac{P(H_0|\boldsymbol{x})}{P(H_1|\boldsymbol{x})} = \frac{0.1152}{1-0.1152} = 0.1301989.$$

Hence $BF_{01} = \frac{0.1301989}{0.0124164} = 10.48604$, and $2\log(BF_{01}) = 4.0009$, showing evidence in favour of H_0.

?

7. The two full-conditional are as follows:
   a) $\mathcal{L}(Z_1, \ldots, Z_k|\text{data}, \theta)$.
   b) $\mathcal{L}(\theta|\text{data}, Z_1, \ldots, Z_k)$

   As far as the full-conditional in a) is concerned, from the "augmentation" method, we have:

$$\mathcal{L}(Z_1, \ldots, Z_k|\text{data}, \theta) = \mathcal{L}(Z_1, \ldots, Z_k|X^+_{m+1}, \ldots, X^+_{m+k}, \theta) = \prod_{j=1}^{k} \mathcal{L}(Z_j|\theta, X^+_{m+j}),$$

   where, for any $j = 1, \ldots, k$, $\mathcal{L}(Z_j|\theta, X^+_{m+j})$ is the distribution of an exponential r.v. bounded to assume values larger than X^+_{m+j}; hence its density is as follows:

$$f_{Z_j}(z_j|\theta, X^+_{m+j}) = \frac{1}{S(X^+_{m+j}; \theta)} f(z_j; \theta) \mathbf{1}_{(X^+_{m+j}, +\infty)}(z_j) \propto \theta e^{-\theta(z_j - X^+_{m+j})} \mathbf{1}_{(X^+_{m+j}, +\infty)}(z_j), \tag{5}$$

   where $f(\cdot; \theta)$ and $S(\cdot; \theta)$ denotes the density and survival functions of an exponential r.v. with failure rate θ, respectively. This distribution is the truncated exponential distr. with failure rate θ over the support $(X^+_{m+j}, +\infty)$, and can be simulated as $X^+_{m+j} + Z$, where $Z \sim \mathit{Exp}(\theta)$.

   We compute the full conditional in b) from Bayes' theorem:

$$\mathcal{L}(\theta|\text{data}, Z_1, \ldots, Z_k) \propto \mathcal{L}(Z_1, \ldots, Z_k, \text{data}|\theta)\mathcal{L}(\theta).$$

   This is the posterior for θ when $X_1, \ldots, X_n, Z_1, \ldots, Z_k$ are from model (1) with conjugate prior $\pi_2 = \mathit{gamma}(\alpha, \beta)$. Hence we have:

$$\mathcal{L}(\theta|\text{data}, Z_1, \ldots, Z_k) = \mathcal{L}(\theta|X_1, \ldots, X_m, Z_1, \ldots, Z_k) \stackrel{d}{=} \mathit{gamma}(\alpha + m + k, \beta + \sum_{i=1}^{m} X_i + \sum_{i=1}^{k} Z_i).$$

BAYESIAN STATISTICS A. Guglielmi 29.01.2019

Properly justify all your answers. Use the indicator function to denote the support of a distr..

Exercise 1 Given parameters K and θ, let the conditional density of the r.v. X_1 be

$$f(x|K,\theta) = \frac{1}{K!}\theta^{K+1}x^K e^{-\theta x}\mathbf{1}_{(0,\infty)}(x), \tag{1}$$

where $K \in \{0, 1, \dots\}$ and $\theta > 0$. As a prior for (K, θ) assume

$$\pi(k,\theta) = a^k(1-a)\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}\mathbf{1}_{\{0,1,\dots\}}(k)\mathbf{1}_{(0,+\infty)}(\theta), \tag{2}$$

where $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}e^{-x}dx$ when $\alpha > 0$.

1. For which values of a and (α, β) (2) defines a proper probability density w.r.t. some measure on $\mathbb{R}^2$?
2. Derive the marginal (prior) density m_{X_1} of X_1. Compute the marginal (prior) mean $E(X_1)$; for which values of hyperparameters the mean is finite?
3. Compute the joint posterior density of (K, θ), given $X_1 = x_1$, using the expression of the marginal prior density m_{X_1}. In addition, find the marginal posterior densities of K and of θ. These distributions are well-known: which distributions are they?
4. Assume $x_1 = 3$, $a = 2/3$, $\alpha = 1$. Test the hypotheses:

   $$H_0\colon \theta \le 1 \text{ vs } H_1\colon \theta > 1,$$

   computing the Bayes factor as the ratio between posterior and prior odds as a function of β. When $\beta = 10$ which hypothesis do you choose? Why?
5. Consider now a sample of size n, such that the conditional distribution of X_i's is given by (1) with prior (2); denotes by $x_1, \dots, x_n$ the observed data. Build a Gibbs sampler for the posterior of (K, θ), writing all the steps in detail. If one of the full-conditionals is not in closed form, suggest how to sample from it.

   (**Hint**: *it could be useful to sample from a Poisson density with an appropriate parameter.*)

Solution of Ex. 1.
From (1) we see that $X_1|K,\theta \sim \text{gamma}(K+1,\theta)$. Moreover, the prior in (2) is the product of two densities, one is the density $\pi(k)$ (wrt the counting measure on $\mathbb{R}$), and the other is the density $\pi(\theta)$ (wrt the Lebesgue measure on $\mathbb{R}$). Hence K and θ are a priori independent and $\pi(k,\theta) = \pi(k)\pi(\theta)$.

1. It is straightforward to see from (2) that $\pi(\theta)$ is the *gamma*(α, β) density when $\alpha, \beta > 0$. As far as $\pi(k)$ is concerned, we observe that:

   (a) $\pi(k) \ge 0$ for all k and $\pi(k) > 0$ for at least one k in the support implies $a > 0$,

   (b) The geometric series $\sum_{k=1}^{\infty} a^k$ converges to $1/(1-a)$ if and only if $-1 \le a < 1$

   By (a) and (b) we need to assume $a \in (0, 1)$.

2. The marginal (prior) density m_{X_1} of X_1 derived from (1)-(2) is as follows, when $x_1 > 0$:

$$m_{X_1}(x_1) = \int_0^{+\infty} \sum_{k=0}^{+\infty} f(x_1|k,\theta)\pi(k,\theta)d\theta = (1-a)\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^{+\infty}\left(\sum_{k=0}^{+\infty}\frac{1}{k!}(ax\theta)^k\right)\theta^{1+\alpha-1}e^{-\theta(x_1+\beta)}d\theta$$
$$= (1-a)\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^{+\infty}\theta^{\alpha+1-1}e^{-(\beta+x_1(1-a))\theta}d\theta = (1-a)\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+1)}{(\beta+x_1(1-a))^{\alpha+1}}$$
$$= (1-a)\frac{\alpha\beta^\alpha}{(\beta+(1-a)x_1)^{\alpha+1}}, \text{ where } a\in(0,1) \text{ and } \alpha,\beta>0.$$

The density $m_{X_1}(x_1) = 0$ if $x_1 < 0$.

To compute $E(X_1)$, instead of integrating wrt the density m_{X_1}, from well-known properties of the conditional mean and the a priori independence between K and θ, we have:

$$E(X_1) = E_\pi(E(X_1|K,\theta)) = E_\pi\left(\frac{K+1}{\theta}\right) = E_\pi\left(E_\pi\left(\frac{K+1}{\theta}|\theta\right)\right) = E_\pi\left(\frac{1}{\theta}E_\pi(K+1|\theta)\right)$$
$$= E_\pi\left(\frac{1}{\theta}\right)E_\pi(K+1) = \frac{\beta}{\alpha-1}(1+E_\pi(K)) = \frac{\beta}{\alpha-1}\left(1+\frac{a}{1-a}\right) = \frac{1}{1-a}\frac{\beta}{\alpha-1}$$

if $\alpha > 1$, given that, in this case, $E_\pi(1/\theta) = \beta/(\alpha-1)$, and that

$$E(K) = (1-a)\sum_0^{+\infty}ka^k = (1-a)a\sum_1^{+\infty}ka^{k-1} = (1-a)a\sum_1^{+\infty}\frac{d}{da}a^k = (1-a)a\frac{d}{da}\sum_1^{+\infty}a^k$$
$$= (1-a)a\frac{d}{da}\left(\frac{1}{1-a}-1\right) = (1-a)a\frac{d}{da}\frac{a}{1-a} = (1-a)a\frac{1}{(1-a)^2} = \frac{a}{1-a},\ a\in(0,1).$$

3. From Bayes's theorem we derive the posterior of (K,θ):

$$\pi(k,\theta|x_1) = \frac{f(x_1|k,\theta)\pi(\kappa,\theta)}{m_{X_1}(x_1)} = \frac{\frac{1}{k!}\theta^{k+1}x_1^k e^{-\theta x_1}a^k(1-a)\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}}{(1-a)\frac{\alpha\beta^\alpha}{(\beta+(1-a)x_1)^{\alpha+1}}}\mathbf{1}_{\{0,1,\dots\}}(k)\mathbf{1}_{(0,+\infty)}(\theta)$$
$$= \frac{(\beta+(1-a)x_1)^{\alpha+1}}{\Gamma(\alpha+1)}\frac{1}{k!}e^{-(x_1+\beta)\theta}(a\theta x_1)^k\theta^{\alpha+1-1}\mathbf{1}_{\{0,1,\dots\}}(k)\mathbf{1}_{(0,+\infty)}(\theta).$$

The marginal posterior density of θ is:

$$\begin{aligned}\pi(\theta|x_1) &= \sum_{k=0}^{\infty}\pi(k,\theta) = \frac{(\beta+(1-a)x_1)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha+1-1}e^{-(x_1+\beta)\theta}\sum_{k=0}^{\infty}\frac{1}{k!}(a\theta x_1)^k = \\ &= \frac{(\beta+(1-a)x_1)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha+1-1}e^{-(x_1+\beta)\theta}e^{a\theta x_1} \\ &= \frac{(\beta+(1-a)x_1)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^{\alpha+1-1}e^{-(\beta+(1-a)x_1)\theta}\mathbf{1}_{(0,+\infty)}(\theta).\end{aligned}$$

If we denote by $\mu = \beta+(1-a)x_1$, we have that $\theta|x_1 \sim gamma(\alpha+1,\mu)$.

The marginal posterior density of K is:

$$\begin{aligned}\pi(k|x_1) &= \int_0^\infty \pi(k,\theta)d\theta = \frac{(\beta+(1-a)x_1)^{\alpha+1}}{\Gamma(\alpha+1)}\frac{1}{k!}(ax_1)^k\int_0^\infty \theta^{\alpha+1+k-1}e^{-(x_1+\beta)\theta}d\theta = \\ &= \frac{(\beta+(1-a)x_1)^{\alpha+1}}{\Gamma(\alpha+1)}\frac{1}{k!}(ax_1)^k\frac{\Gamma(\alpha+1+k)}{(x_1+\beta)^{\alpha+1+k}} \\ &= \frac{\Gamma(\alpha+1+k)}{\Gamma(\alpha+1)k!}\left(\frac{\beta+(1-a)x_1}{x_1+\beta}\right)^{\alpha+1}\left(\frac{ax_1}{x_1+\beta}\right)^k, \quad k=0,1,\dots .\end{aligned}$$

This is the density of the negative binomial distribution with parameters $r = \alpha + 1$ (*generalized* since r is not an integer) and $p = \frac{\beta+(1-a)x_1}{x_1+\beta}$.

4. First of all, note that, if $\alpha = 1$, then a priori $\theta \sim Exp(\beta)$ with distribution function $F_\pi(\theta) = 1 - \mathrm{e}^{-\beta\theta}$ when $\theta > 0$. A posteriori $\theta|x_1 \sim gamma(2, \mu = \beta + (1-a)x_1)$ con funzione di ripartizione:

$$F_\pi(\theta|x_1) = \int_0^\theta \mu^2 u \mathrm{e}^{-\mu u} du = \mu^2 \left[u \left(-\frac{1}{\mu} \mathrm{e}^{-\mu u} \right) \right]_0^\theta - \mu^2 \int_0^\theta -\frac{1}{\mu} \mathrm{e}^{-\mu u} du = -\theta\mu \mathrm{e}^{-\mu\theta} + 1 - \mathrm{e}^{-\mu\theta}$$
$$= 1 - (1 + \mu\theta)\mathrm{e}^{-\mu\theta}, \; \theta > 0.$$

In particular, $x_1 = 3$, $a = 2/3$, $\alpha = 1$, $\beta = 10$ yields $\mu = \beta + 1 = 11$. Prior odds is equal to

$$\frac{F_\pi(1)}{1 - F_\pi(1)} = \frac{1 - \mathrm{e}^{-\beta}}{\mathrm{e}^{-\beta}} \simeq 22025.47$$

The Bayes factor is:

$$\mathrm{BF}_{01} = \frac{\frac{\mathbb{P}(\theta \leq 1|x_1)}{\mathbb{P}(\theta > 1|x_1)}}{\frac{\mathbb{P}(\theta \leq 1)}{\mathbb{P}(\theta > 1)}} = \frac{\frac{F_\pi(1|x_1)}{1 - F_\pi(1|x_1)}}{\frac{F_\pi(1)}{1 - F_\pi(1)}} = \frac{\frac{1-(1+\mu)\mathrm{e}^{-\mu}}{(1+\mu)\mathrm{e}^{-\mu}}}{\frac{1-\mathrm{e}^{-\beta}}{\mathrm{e}^{-\beta}}} = \frac{1 - (\beta+2)\mathrm{e}^{\beta+1}}{\mathrm{e}^{-1}(\beta+2)(1-\mathrm{e}^{-\beta})} \simeq \frac{4988.512}{22025.47} \simeq 0.2265.$$

Since $2\log(BF_{01}) = -2.970021 \in (-5, -2)$, we conclude that there is evidence in favour of $H_1 : \theta > 1$.

5. Let $\boldsymbol{x} = (x_1, \ldots, x_n)$. If these data are assumed iid from (1) with prior (2), then the posterior distribution of (K, θ) has density:

$$\pi(k, \theta|\underline{x}) \propto \prod_{i=1}^n \{f(x_i|k,\theta)\}\pi(k,\theta)$$
$$= \left(\frac{1}{k!}\right)^n \theta^{n(k+1)} \left(\prod_1^n x_1\right)^k \mathrm{e}^{-\theta\sum_{i=1}^n x_i} a^k (1-a) \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \mathrm{e}^{-\beta\theta} \mathbf{1}_{\{0,1,\ldots\}}(k) \mathbf{1}_{(0,+\infty)}(\theta).$$

To build a Gibbs sampler the two *full conditionals* $\pi(\theta|k, \underline{x})$ and $\pi(k|\theta, \underline{x})$ have to be derived.

- $\pi(\theta|k, \underline{x})$:
$$\pi(\theta|k, \underline{x}) \propto \theta^{\alpha+n(k+1)-1} \mathrm{e}^{-\theta(\beta+\sum_{i=1}^n x_i)} \mathbf{1}_{(0,+\infty)}(\theta).$$
This is the kernel of a $gamma(\alpha + n(k+1), \beta + \sum_{i=1}^n x_i)$ distribution.
- $p(k|\theta, \underline{x})$:
$$p(k|\theta, \underline{x}) \propto \left(\frac{1}{k!}\right)^n \left(a\theta^n \prod_{i=1}^n x_i\right)^k \mathbf{1}_{\{0,1,\ldots\}}(k),$$
which is not a well-known distribution.

Hence, to simulate from $p(k|\theta, \underline{x})$ we need a Metropolis step.

In particolar, the proposal is the density $q(k)$ of the Poisson$(a\theta^n \prod_{i=1}^n x_i)$. Let $(k^{(g)}, \theta^g)$, $g \in \{1, \ldots, G\}$ be the current state of the chain. Using notation $\theta := \theta^{(g)}$, let now $k' \sim$Poisson$(a\theta^n \prod_{i=1}^n x_i)$ be the proposal for the new value of k. Then $k^{(g+1)}$ is equal to k' with probability $\min\{1, \alpha\}$, where

$$\alpha = \frac{\pi(k'|\theta, \underline{x})q(k^{(g)})}{\pi(k^{(g)}|\theta, \underline{x})q(k')} = \frac{\left(\frac{1}{k'!}\right)^n \left(a\theta^n \prod_{i=1}^n x_i\right)^{k'} \mathrm{e}^{-(a\theta^n \prod_{i=1}^n x_i)} \left(a\theta^n \prod_{i=1}^n x_i\right)^{k^{(g)}} \left(\frac{1}{k^{(g)}!}\right)}{\left(\frac{1}{k^{(g)}!}\right)^n \left(a\theta^n \prod_{i=1}^n x_i\right)^{k^{(g)}} \mathrm{e}^{-(a\theta^n \prod_{i=1}^n x_i)} \left(a\theta^n \prod_{i=1}^n x_i\right)^{k'} \left(\frac{1}{k'!}\right)} = \left(\frac{k^{(g)}!}{k'!}\right)^{n-1}.$$

Exercise 2 Describe the formal approach to model choice in the Bayesian approach, in particular when there are K possible Bayesian models $M_1, \dots, M_K$. Specify the prior probability on the set of models $\{M_1, \dots, M_K\}$, the likelihood for data $\boldsymbol{y}$ and the prior distribution for any model M_j, $j = 1, \dots, K$. Which (model or) models are the *best* ones?

Solution of Ex 2.
Let $\mathbb{P}(m = j)$ be the prior probability of model M_j, $j = 1, \dots, K$; for instance, we typically choose it as $1/K$.
Let us denote by $f(\boldsymbol{y}|\theta_j, M_j)$ the likelihood under model M_j, and the prior $\pi(\theta_j|M_j) = \pi(\theta_j|m = j)$ for the parameter vector θ_j.

Then, for any model M_j, compute the posterior of parameter θ_j according Bayes' theorem:

$$\pi(\theta_j|\boldsymbol{y}, M_j) = \frac{f(\boldsymbol{y}|\theta_j, M_j)\pi(\theta_j|M_j)}{m(\boldsymbol{y}|M_j)},$$

where $m(\boldsymbol{y}|M_j) = \int f(\boldsymbol{y}|\theta_j, M_j)\pi(\theta_j|M_j)d\theta_j$ is the (prior) marginal distribution of data $\boldsymbol{y}$ under model M_j.

Compute also the posterior probability masses of $M_1, \dots, M_K$:

$$\mathbb{P}(m = j|\boldsymbol{y}) = \frac{m(\boldsymbol{y}|M_j)\mathbb{P}(m = j)}{m(\boldsymbol{y})}, \quad j = 1, \dots, K,$$

with $m(\boldsymbol{y}) = \sum_{j=1}^{K} m(\boldsymbol{y}|M_j)\mathbb{P}(m = j)$.

Finally, choose the model with the highest $\mathbb{P}(m = j|\boldsymbol{y})$ as the *best* one. Alternatively, choose a few models (e.g. 2-3 models) with the highest posterior probabilities, and then compare these few models via predictive goodness-of-fit criteria.

■

BAYESIAN STATISTICS A. Guglielmi 05.07.2018

Properly justify all your answers.

Exercise 1 Consider the following autoregressive model of order 1 for a sequence of r.v.'s $\eta_1, \eta_2, \ldots, \eta_T$, where

(1) $$\eta_1 \sim \mathcal{N}(0,1)$$

(2) $$\eta_t = \psi\eta_{t-1} + \varepsilon_t, \quad \varepsilon_t|\psi \overset{iid}{\sim} \mathcal{N}(0, 1-\psi^2), \quad t = 2, 3, \ldots, T,$$

with $-1 < \psi < 1$, and η_1 and $\{\varepsilon_t\}$ are independent. Consider also a standard Gaussian r.v. ζ, independent from $\{\eta_t, t = 1, 2, \ldots, T\}$ and define

(3) $$p_t = \mathrm{e}^{-\frac{\zeta^2+\eta_t^2}{2\alpha}}, \quad t = 1, 2, \ldots, T, \quad \text{with } \alpha > 0.$$

1. Derive $\mathrm{E}(\eta_t|\psi)$, $\mathrm{Var}(\eta_t|\psi)$ and $\mathrm{Cov}(\eta_{t-1}, \eta_t|\psi)$ for any $t = 2, 3, \ldots, T$ (mean and variance of η_1 appear in (1)). Derive the conditional marginal distribution of any η_t, for $t = 1, 2, \ldots, T$, given ψ. Does this distribution depend on ψ? Make an explicit comment.

Now assume that p_1, p_2 are random parameters with a prior distribution defined as in (3) ($T = 2$), and that ψ is a priori distributed according to the uniform on $(-1, 1)$. There is a group of n cancer patients who respond to a treatment with probability p_1 and p_2 at time $t = 1$ and $t = 2$, respectively. Let $X_{1t}, X_{2t}, \ldots, X_{nt}$ be r.v.'s equal to 1 if the corresponding patient responds to the treatment at time t, and 0 otherwise; assume that, conditionally to $(\psi, \zeta, \eta_1, \eta_2)$, $\boldsymbol{X}_1 := (X_{11}, X_{21}, \ldots, X_{n1})$ and $\boldsymbol{X}_2 := (X_{12}, X_{22}, \ldots, X_{n2})$ are independent, each X_{jt} being iid according to a Bernoulli distribution with success probability equal to p_t, for $t = 1, 2$. Assume that ζ and (ψ, η_1, η_2) are a priori independent. We also fix $\alpha = 1$ in points 2. and 3. below to make calculations simpler.

2. Write the expression of the posterior distribution of $(\psi, \zeta, \eta_1, \eta_2)$, given $\boldsymbol{X}_1, \boldsymbol{X}_2$, up to a constant.
3. Outline how you would build an hybrid Gibbs sampler algorithm for the posterior distribution of $(\psi, \zeta, \eta_1, \eta_2)$.

Assume $\alpha > 0$ known for the rest of the text.

4. Prove that the marginal distribution of any p_t, for $t = 1, 2, \ldots, T$, is *beta*$(\alpha, 1)$.
   (**Hint**: *first derive the distribution of* $Y_t := \zeta^2 + \eta_t^2$.)

Consider $\{p_{l,t}, l = 1, 2, 3, \ldots\}$ iid copies of p_t, as in (3) and define, for $t = 1, 2$,

(4) $$w_{1,t} = 1 - p_{1,t}$$

(5) $$w_{l,t} = (1 - p_{l,t})p_{1,t}p_{2,t} \cdots p_{l-1,t}, \quad l = 2, 3, \ldots.$$

Let $\theta_l \overset{iid}{\sim} P_0$ for $l = 1, 2, \ldots$, with $\{\theta_l, l = 1, 2, 3, \ldots\}$ and $\{p_{l,t}, l = 1, 2, 3, \ldots, t = 1, 2\}$ independent, where P_0 is some non-atomic probability measure on $\mathbb{R}$. Define two r.p.m.'s on $\mathbb{R}$ as

(6) $$G_t = \sum_{l=1}^{+\infty} w_{l,t}\delta_{\theta_l}, \quad t = 1, 2.$$

5. Compute $\mathrm{E}(G_t(A))$, $\mathrm{Var}(G_t(A))$, for $t = 1, 2$, where A is an interval in $\mathbb{R}$.

Solution of Ex 1.

1. Of course, by (1), $\eta_1 \sim \mathcal{N}(0,1)$. Consequently, conditioning to ψ, $\eta_2 = \psi\eta_1 + \varepsilon_2$ is a linear combination of two independent Gaussian r.v's and hence it is a Gaussian r.v., and so on, any $\eta_t = \psi\eta_{t-1} + \varepsilon_t$ is a Gaussian r.v..

   Moreover, $\mathrm{E}(\eta_1|\psi) = \mathrm{E}(\eta_1) = 0$, so that $\mathrm{E}(\eta_2|\psi) = \mathrm{E}[\mathrm{E}(\eta_2|\eta_1,\psi)] = \mathrm{E}[\psi\eta_1|\psi] = \psi\,\mathrm{E}[\eta_1|\psi] = 0$, and iteratively $\mathrm{E}(\eta_t|\psi) = \mathrm{E}[\mathrm{E}(\eta_t|\eta_{t-1},\psi)] = \psi\,\mathrm{E}[\eta_{t-1}|\psi] = 0$ for all t.

   Similarly, $\mathrm{Var}(\eta_1|\psi) = \mathrm{Var}(\eta_1) = 1$, and, because of independence between ε_t and η_{t-1}, $\mathrm{Var}(\eta_t|\psi) = \mathrm{Var}(\psi\eta_{t-1} + \varepsilon_t|\psi) = \psi^2\,\mathrm{Var}(\eta_{t-1}|\psi) + \mathrm{Var}(\varepsilon_t|\psi) = \psi^2 \times 1 + 1 - \psi^2 = 1$ for any t.

   Summing up, the conditional marginal distribution of $\eta_t|\psi$ is $\mathcal{N}(0,1)$ which does not depend on ψ, i.e. the marginal distribution of each η_t is $\mathcal{N}(0,1)$.

   It is also straightforward to prove that, conditionally to ψ, for any $t = 2,3,\ldots,T$,

$$\mathrm{Cov}(\eta_{t-1},\eta_t|\psi) = \mathrm{E}(\eta_{t-1}\eta_t|\psi) = \mathrm{E}[\mathrm{E}(\eta_{t-1}\eta_t|\eta_{t-1},\psi)] = \mathrm{E}[\eta_{t-1}\psi\eta_{t-1}|\psi] = \psi\,\mathrm{Var}(\eta_{t-1}|\psi) = \psi.$$

2. We have

$$\begin{aligned}\mathcal{L}(\psi,\zeta,\eta_1,\eta_2|\boldsymbol{X}_1,\boldsymbol{X}_2) &\propto \mathcal{L}(\boldsymbol{X}_1,\boldsymbol{X}_2|\psi,\zeta,\eta_1,\eta_2) \times \pi(\zeta) \times \pi(\eta_1) \times \pi(\eta_2|\eta_1,\psi) \times \pi(\psi)\\ &\propto p_1^{\sum_1^n x_{i1}}(1-p_1)^{n-\sum_1^n x_{i1}} p_2^{\sum_1^n x_{i2}}(1-p_2)^{n-\sum_1^n x_{i2}} \times \pi(\zeta) \times \pi(\eta_1) \times \pi(\eta_2|\eta_1,\psi) \times \pi(\psi).\end{aligned}$$

   Note that, since $\eta_2|\eta_1,\psi \sim \mathcal{N}(\psi\eta_1, 1-\psi^2)$ and $\eta_1 \sim \mathcal{N}(0,1)$, then $(\eta_1,\eta_2|\psi)^* \sim \mathcal{N}_2((0,0),\begin{pmatrix}1 & \psi\\ \psi & 1\end{pmatrix})$, where the $*$ denotes the transpose of a row vector, i.e. a column vector.

   We have

$$\begin{aligned}\mathcal{L}(\psi,\zeta,\eta_1,\eta_2|\boldsymbol{X}_1,\boldsymbol{X}_2) \propto\ & \mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)})^{n-\sum_1^n x_{i1}}\\ &\times \mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)})^{n-\sum_1^n x_{i2}}\\ &\times \mathcal{N}(\eta_1;0,1) \times \mathcal{N}(\eta_2;\psi\eta_1,1-\psi^2) \times \mathcal{N}(\zeta;0,1) \times \mathbf{1}_{(-1,1)}(\psi).\end{aligned}$$

3. The Gibbs sampler iteratively samples from the full-conditionals of the 4 parameters

   (a) $[\psi|rest] \propto \mathcal{N}(\eta_2;\psi\eta_1,1-\psi^2) \times \mathbf{1}_{(-1,1)}(\psi) \propto \mathrm{e}^{-\frac{1}{2}\log(1-\psi^2)-\frac{(\eta_2-\psi\eta_1)^2}{2(1-\psi^2)}}\mathbf{1}_{(-1,1)}(\psi)$

   (b) $[\eta_2|rest] \propto \mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)})^{n-\sum_1^n x_{i2}} \times \mathcal{N}(\eta_2;\psi\eta_1,1-\psi^2)$
   $\propto \mathrm{e}^{-(\sum_1^n x_{i2}/2)(\eta_2^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)})^{n-\sum_1^n x_{i2}} \times \mathrm{e}^{\frac{(\eta_2-\psi\eta_1)^2}{2(1-\psi)^2}}$

   (c) $[\eta_1|rest] \propto \mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)})^{n-\sum_1^n x_{i1}} \times \mathcal{N}(\eta_1;0,1)$
   $\propto \mathrm{e}^{-((1+\sum_1^n x_{i1})/2)(\eta_1^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)})^{n-\sum_1^n x_{i1}}$

   (d)

$$\begin{aligned}[\zeta|rest] \propto\ & \mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)})^{n-\sum_1^n x_{i1}}\\ &\times \mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)}(1-\mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)})^{n-\sum_1^n x_{i2}} \times \mathcal{N}(\zeta;0,1)\\ &\propto \mathrm{e}^{-\zeta^2(1+\sum_1^n x_{i1}+\sum_1^n x_{i2})/2}(1-\mathrm{e}^{-(\sum_1^n x_{i1}/2)(\zeta^2+\eta_1^2)})^{n-\sum_1^n x_{i1}}\\ &\qquad \times (1-\mathrm{e}^{-(\sum_1^n x_{i2}/2)(\zeta^2+\eta_2^2)})^{n-\sum_1^n x_{i2}}\end{aligned}$$

4. For any $t = 1,2,\ldots,T$, $Y_t = \zeta^2 + \eta_t^2$ is the sum of the squares of two independent Gaussian r.v.'s and hence its distribution is $\chi^2(2) = gamma(\frac{2}{2},\frac{1}{2})$.

Alternatively, we could remember that, if $\zeta \sim \mathcal{N}(0,1)$, then, for $s > 0$,

$$F_{\zeta^2}(s) = \mathbb{P}(\zeta^2 \le s) = \mathbb{P}(-\sqrt{s} \le \zeta \le \sqrt{s}) = \Phi(\sqrt{s}) - \Phi(-\sqrt{s}),$$

so that the density of ζ^2 has the following expression for $s > 0$:

$$f_{\zeta^2}(s) = \frac{1}{\sqrt{s}}\varphi(\sqrt{s}) = \frac{1}{\sqrt{s}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-s/2} = \frac{1}{\sqrt{2\pi}}s^{\frac{1}{2}-1}\mathrm{e}^{-s/2}\mathbf{1}_{(0,+\infty)(s)} = gamma(\frac{1}{2},\frac{1}{2}) = \chi^2(1).$$

Consequently, Y_t, which is the sum of 2 independent $\chi^2(1)$ r.v.'s, has $\chi^2(2)$-distribution, with density $f(s) = \frac{1}{2}\frac{1}{\Gamma(1)}s^{1-1}\mathrm{e}^{-\frac{s}{2}}\mathbf{1}_{(0,+\infty)}(s) = \frac{1}{2}\mathrm{e}^{-\frac{s}{2}}\mathbf{1}_{(0,+\infty)}(s)$.

It is straightforward to prove that, for $0 < z < 1$,

$$F_{p_t}(z) = \mathbb{P}(p_t \le z) = \mathbb{P}(\mathrm{e}^{-\frac{\zeta^2+\eta_t^2}{2\alpha}} \le z) = \mathbb{P}(-\frac{\zeta^2+\eta_t^2}{2\alpha} \le \log(z)) = \mathbb{P}(Y_t \ge -2\alpha\log(z)) = 1 - F_{Y_t}(-2\alpha\log(z)),$$

so that for $0 < z < 1$,

$$f_{p_t}(z) = -f_{Y_t}(-2\alpha\log(z))(-2\alpha)\frac{1}{z} = \frac{2\alpha}{z}\frac{1}{2}\mathrm{e}^{\frac{1}{2}2\alpha\log(z)} = \alpha z^{\alpha-1}\mathbf{1}_{(0,1)}(z).$$

This shows that $p_t \sim \ beta(\alpha,1)$.

5. We have that, marginally to the rest, $\{p_{l,t}, l = 1,2,3,\ldots\}$ iid copies of $beta(\alpha,1)$ random variables, and that, for any t, G_t is defined according the stick-breaking construction, i.e. marginally $G_t \sim DP(\alpha, P_0)$ for any t. Hence $G_t(A) \sim beta(\alpha P_0(A), \alpha(1-P_0(A)))$ for any interval A, so that

$$\mathrm{E}(G_t(A) = \frac{\alpha P_0(A)}{\alpha P_0(A) + \alpha(1-P_0(A))} = P_0(A),$$

and

$$\begin{aligned}\mathrm{Var}(G_t(A)) &= \frac{\alpha P_0(A) \times \alpha(1-P_0(A))}{(\alpha P_0(A) + \alpha(1-P_0(A)))^2 \times (\alpha P_0(A) + \alpha(1-P_0(A)) + 1)} \\ &= \frac{\alpha^2 P_0(A)(1-P_0(A))}{\alpha^2(\alpha+1)} = \frac{P_0(A)(1-P_0(A))}{\alpha+1}\end{aligned}$$

for both $t = 1,2$.

The construction of G_1, G_2 is taken from
DeYoreo M. and Kottas A. (2018). Modeling for Dynamic Ordinal Regression Relationships: An Application to Estimating Maturity of Rockfish in California. *JASA*
∎

BAYESIAN STATISTICS A. Guglielmi, I. Bianchini 13.02.2018

Properly justify all your answers.

Exercise 1 Consider the following model for count data $(y_1, y_2, \ldots, y_n)$:

$$Y_1, Y_2, \ldots, Y_{A_1} | \lambda_1, A_1 \overset{iid}{\sim} Poisson(\lambda_1), \qquad Y_{A_1+1}, \ldots, Y_n | \lambda_2, A_1 \overset{iid}{\sim} Poisson(\lambda_2)$$

where $A_1 \in \{1, 2, \ldots, n-1\}$ is called *change point* and the two samples are independent, conditional on $A_1, \lambda_1, \lambda_2$. Denote by $A_2 = n - A_1$ the size of the second sample $Y_{A_1+1}, \ldots, Y_n$.

First, consider A_1 fixed and equal to $a_1 \in \{1, 2, \ldots, n-1\}$.

1. Verify that the moment generating function $M(t; \lambda)$ of $X \sim Poisson(\lambda)$ with density $f_X(k; \lambda) = e^{-\lambda}\frac{\lambda^k}{k!}\mathbf{1}_{0,1,2,\ldots}(k)$ for $\lambda > 0$ is
$$M(t; \lambda) = \mathbb{E}\left(e^{tX}\right) = \exp\left(\lambda(e^t - 1)\right) \quad t \in \mathbb{R}.$$
2. Write the likelihood of $(y_1, y_2, \ldots, y_{a_1}, y_{a_1+1}, \ldots, y_n)$.
3. Find the conjugate prior for (λ_1, λ_2) such that $\pi(\lambda_1, \lambda_2) = \pi_1(\lambda_1) \times \pi_2(\lambda_2)$ and derive the hyperparameters of the posterior distribution.

   (Hint: *the marginal priors* π_1 *and* π_2 *can be parameterized by parameters* (α_j, β_j), $j = 1, 2$, *respectively. Denote the posterior hyperparameters as* $(\alpha_{1a_1}, \beta_{1a_1}, \alpha_{2a_2}, \beta_{2a_2})$.)
4. In order to set the values for (α_j, β_j), $j = 1, 2$, we ask to an expert: she thinks that, before the change point, marginal prior values are such that $\mathrm{E}(Y_1) = 4$, $\mathrm{Var}(Y_1) = 8$, and, after the change point, $\mathrm{E}(Y_n) = 8$, $\mathrm{Var}(Y_n) = 16$. Compute the prior hyperparameters and derive the corresponding posterior hyperparameters.

Now, suppose that A_1 is random: in particular, consider A_1 uniformly distributed on $\{1, 2, \ldots, n-1\}$ and independent on (λ_1, λ_2). Here (λ_1, λ_2) has the prior derived at points **??** and **??**.

5. Check that, in this case, the likelihood has the same analytic form as at point **??**. Then determine the marginal posterior of A_1, up to a normalizing constant, denoted by K.

   (Hint: *first derive the joint posterior distribution of* $(A_1, \lambda_1, \lambda_2)$ *and then obtain the marginal posterior of* A_1.)
6. Assume here $n = 4$ and $(y_1, y_2, y_3, y_4) = (0, 0, 2, 0)$. Test the hypotheses $H_0 : A_1 \leq 2$ vs $H_1 : A_1 > 2$ using the Bayes factor, making explicit your conclusion.

Exercise 2 Compute the marginal distribution of a sample $(X_1, X_2, \ldots, X_n)$ from a Dirichlet process P with parameters (α, P_0), where $\alpha > 0$ is the total mass and P_0 is centering probability measure. Provide the proof.

Solution of Ex 1.

1. We have:
$$M(t; \lambda) = \sum_{k=1}^{+\infty} e^{tk} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{+\infty}\frac{(e^t\lambda)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}, \; t \in \mathbb{R}.$$

2. For count data $(y_1, y_2, \ldots, y_n)$, the likelihood here is given by

$$L(\lambda_1, \lambda_2; y_1, \ldots, y_{a_1}, y_{a_1+1}, \ldots, y_n) = \prod_{i=1}^{a_1} f_{Y_i}(y_i; \lambda_1) \prod_{j=a_1+1}^{n} f_{Y_j}(y_j; \lambda_2)$$
$$\propto \lambda_1^{\sum_{i=1}^{a_1} y_i} \mathrm{e}^{-\lambda_1 a_1} \lambda_2^{\sum_{j=a_1+1}^{n} y_j} \mathrm{e}^{-\lambda_2 a_2}, \text{ where } \lambda_1 > 0 \text{ and } \lambda_2 > 0.$$

3. By inspecting the analytic form of the likelihood, it is clear that the conjugate prior of the product form is such that the marginal priors are both gamma laws. Thus, $\pi_1(\lambda_1) = gamma(\alpha_1, \beta_1)$ and $\pi_2(\lambda_2) = gamma(\alpha_2, \beta_2)$. We obtain:

$$\pi(\lambda_1, \lambda_2 | y_1, \ldots, y_{a_1}, y_{a_1+1}, \ldots, y_n) \propto \lambda_1^{\sum_{i=1}^{a_1} y_i} \mathrm{e}^{-\lambda_1 a_1} \lambda_2^{\sum_{j=a_1+1}^{n} y_j} \mathrm{e}^{-\lambda_2 a_2}$$
$$\times \lambda_1^{\alpha_1 - 1} \mathrm{e}^{-\beta_1 \lambda_1} \lambda_2^{\alpha_2 - 1} \mathrm{e}^{-\beta_2 \lambda_2} \mathbf{1}_{(0,+\infty)}(\lambda_1) \mathbf{1}_{(0,+\infty)}(\lambda_2)$$
$$\propto \lambda_1^{\alpha_1 + \sum_{i=1}^{a_1} y_i - 1} \mathrm{e}^{-(\beta_1 + a_1)\lambda_1} \lambda_2^{\alpha_2 + \sum_{j=a_1+1}^{n} y_j - 1} \mathrm{e}^{-(\beta_2 + a_2)\lambda_2} \mathbf{1}_{(0,+\infty)}(\lambda_1) \mathbf{1}_{(0,+\infty)}(\lambda_2)$$

Hence, posterior hyperparameters are given by:

$$\alpha_{1a_1} = \alpha_1 + \sum_{i=1}^{a_1} y_i, \qquad \beta_{1a_1} = \beta_1 + a_1$$

$$\alpha_{2a_2} = \alpha_2 + \sum_{j=a_1+1}^{n} y_j, \qquad \beta_{2a_2} = \beta_2 + a_2.$$

4. In order to set values for (α_j, β_j), $j = 1, 2$, we can either specify the marginal distribution

$$m(Y) = \int_0^{+\infty} Poisson(Y; \lambda) gamma(d\lambda; \alpha, \beta)$$

for Y_1 and Y_n, or directly compute marginal mean and variance, as follows:

$$\mathrm{E}(Y_1) = \mathrm{E}(\mathrm{E}(Y_1 | \lambda_1)) = \mathrm{E}(\lambda_1) = \frac{\alpha_1}{\beta_1},$$

$$\mathrm{Var}(Y_1) = \mathrm{E}(\mathrm{Var}(Y_1 | \lambda_1)) + \mathrm{Var}(\mathrm{E}(Y_1 | \lambda_1)) = \mathrm{E}(\lambda_1) + \mathrm{Var}(\lambda_1) = \frac{\alpha_1(\beta_1 + 1)}{\beta_1^2},$$

and a similar argument can be applied for Y_n. Note that, by 1.,

$$\mathrm{E}(Y_1 | \lambda_1) = M'(t, \lambda_1)|_{t=0} = \lambda_1 \mathrm{e}^t \mathrm{e}^{\lambda_1(\mathrm{e}^t - 1)}|_{t=0} = \lambda_1,$$

and

$$\mathrm{E}(Y_1^2 | \lambda_1) = M''(t, \lambda_1)|_{t=0} = \lambda_1 \mathrm{e}^t \mathrm{e}^{\lambda_1(\mathrm{e}^t - 1)}(1 + \lambda_1 \mathrm{e}^t)|_{t=0} = \lambda_1(1 + \lambda_1),$$

so that $\mathrm{Var}(Y_1 | \lambda_1) = \lambda_1$, and similarly for Y_2.

Thus, we have to solve two linear systems:

$$\frac{\alpha_1}{\beta_1} = 4, \qquad \frac{\alpha_1(\beta_1 + 1)}{\beta_1^2} = 8$$

$$\frac{\alpha_2}{\beta_2} = 8, \qquad \frac{\alpha_2(\beta_2 + 1)}{\beta_2^2} = 16$$

leading to $(\alpha_1, \beta_1) = (4, 1)$ and $(\alpha_2, \beta_2) = (8, 1)$. Posterior hyperparameters are: $\alpha_{1a_1} = 4 + \sum_{i=1}^{a_1} y_i$, $\beta_{1a_1} = 1 + a_1$, $\alpha_{2a_2} = 8 + \sum_{j=a_1+1}^{n} y_j$, $\beta_{2a_2} = 1 + a_2$.

5. First of all, observe that the likelihood for parameter $(A_1, \lambda_1, \lambda_2)$ has the same (formal) expression as before, up to the support of the marginal prior of A_1, since $\pi(a_1) = \frac{1}{n-1}$ =const for $a_i = 1, \ldots, n-1$:

$$L(a_1, \lambda_1, \lambda_2; y_1, \ldots, y_{a_1}, y_{a_1+1}, \ldots, y_n) \propto \lambda_1^{\sum_{i=1}^{a_1} y_i} e^{-\lambda_1 a_1} \lambda_2^{\sum_{j=a_1+1}^{n} y_j} e^{-\lambda_2 a_2},\ a_1 = 1, 2, \ldots, n-1,\ \lambda_1, \lambda_2 > 0.$$

Posterior distribution of $(A_1, \lambda_1, \lambda_2)$ is proportional to the product of the likelihood and the prior, so that the marginal posterior distribution of A_1 can be obtained integrating out λ_1 and λ_2 from the joint posterior $\pi(\lambda_1, \lambda_2, A_1|data)$:

$$\mathbb{P}(A_1 = a_1|data) = \pi(a_1|data) = \int_0^{+\infty} \int_0^{+\infty} \pi(A_1, d\lambda_1, d\lambda_2|data).$$

Here,

$$\pi(a_1, d\lambda_1, d\lambda_2|data) \propto \lambda_1^{\alpha_1+\sum_{i=1}^{a_1} y_i - 1} e^{-\lambda_1(\beta_1+a_1)} \lambda_2^{\alpha_2+\sum_{j=a_1+1}^{n} y_j - 1} e^{-\lambda_2(\beta_2+a_2)} \frac{1}{n-1} \mathbf{1}_{\{1,2,\ldots,n-1\}}(a_1) \mathbf{1}_{(0,+\infty)^2}(\lambda_1, \lambda_2)$$

Thus,

$$\begin{aligned}
\pi(a_1|data) &= \int_0^{+\infty} \int_0^{+\infty} \pi(d\lambda_1, d\lambda_2, A_1|data) \\
&\propto \int_0^{+\infty} \int_0^{+\infty} \lambda_1^{\alpha_1+\sum_{i=1}^{a_1} y_i - 1} e^{-\lambda_1(\beta_1+a_1)} \lambda_2^{\alpha_2+\sum_{j=a_1+1}^{n} y_j - 1} e^{-\lambda_2(\beta_2+a_2)} \frac{1}{n-1} \mathbb{I}_{\{1,2,\ldots,n-1\}}(a_1) d\lambda_1 d\lambda_2 \\
&\propto \frac{\Gamma(\alpha_1 + \sum_{i=1}^{a_1} y_i)}{(\beta_1 + a_1)^{\alpha_1+\sum_{i=1}^{a_1} y_i}} \frac{\Gamma(\alpha_2 + \sum_{j=a_1+1}^{n} y_j)}{(\beta_2 + a_2)^{\alpha_2+\sum_{j=a_1+1}^{n} y_j}} \mathbb{I}_{\{1,2,\ldots,n-1\}}(a_1) \\
&= \frac{1}{K} \frac{\Gamma(\alpha_1 + \sum_{i=1}^{a_1} y_i)}{(\beta_1 + a_1)^{\alpha_1+\sum_{i=1}^{a_1} y_i}} \frac{\Gamma(\alpha_2 + \sum_{j=a_1+1}^{n} y_j)}{(\beta_2 + a_2)^{\alpha_2+\sum_{j=a_1+1}^{n} y_j}} \mathbb{I}_{\{1,2,\ldots,n-1\}}(a_1).
\end{aligned}$$

6. The Bayes Factor coincides with the ratio between posterior and prior odds;

$$BF_{01} = \frac{\dfrac{\mathbb{P}(H_0|data)}{1 - \mathbb{P}(H_0|data)}}{\dfrac{\mathbb{P}(H_0)}{1 - \mathbb{P}(H_0)}} = \frac{\dfrac{\mathbb{P}(A_1 \le 2|data)}{\mathbb{P}(A_1 > 2|data)}}{\dfrac{\mathbb{P}(A_1 \le 2)}{\mathbb{P}(A_1 > 2)}}.$$

First of all, note that the prior odds are such that:

$$\frac{\mathbb{P}(A_1 \le 2)}{\mathbb{P}(A_1 > 2)} = \frac{\mathbb{P}(A_1 \in \{1, 2\})}{\mathbb{P}(A_1 \in \{3\})} = \frac{\frac{2}{3}}{\frac{1}{3}} = 2\,.$$

Moreover, we have that

$$\mathbb{P}(A_1 \in \{3\}|data) = \frac{1}{K} \left(\frac{\Gamma(\alpha_1 + \sum_{i=1}^{3} y_i)}{(\beta_1 + 3)^{\alpha_1+\sum_{i=1}^{3} y_i}} \frac{\Gamma(\alpha_2 + y_4)}{(\beta_2 + 1)^{\alpha_2+y_4}} \right)$$

and

$$\mathbb{P}(A_1 \in \{1, 2\}|data) = \frac{1}{K} \left(\frac{\Gamma(\alpha_1 + y_1)}{(\beta_1 + 1)^{\alpha_1+y_1}} \frac{\Gamma(\alpha_2 + \sum_{i=2}^{4} y_i)}{(\beta_2 + 3)^{\alpha_2+\sum_{i=2}^{4} y_i}} + \frac{\Gamma(\alpha_1 + \sum_{i=1}^{2} y_i)}{(\beta_1 + 2)^{\alpha_1+\sum_{i=1}^{2} y_i}} \frac{\Gamma(\alpha_2 + \sum_{i=3}^{4} y_i)}{(\beta_2 + 2)^{\alpha_2+\sum_{i=3}^{4} y_i}} \right).$$

Note that the knowledge of the normalizing constant K is not needed, since it simplifies when computing the Bayes Factor.

When $n = 4$, $(y_1, y_2, y_3, y_4) = (0, 0, 2, 0)$ the BF is equal to

$$\begin{aligned}
\text{posterior odds} &= \left(\frac{\Gamma(4+0)\Gamma(8+2)}{(1+1)^4(1+3)^{10}} + \frac{\Gamma(4+0)\Gamma(8+2)}{(1+2)^4(1+2)^{10}}\right) \times \left(\frac{\Gamma(4+2)\Gamma(8+0)}{(1+3)^6(1+1)^8}\right)^{-1} \\
&= \frac{18}{5}\left(\frac{1}{2}\right)^4 \left(1 + \frac{1}{(3/2)^4(3/4)^{10}}\right) \simeq \frac{0.5850}{0.5768} = 1.0142,
\end{aligned}$$

so that

$$BF_{01} = \frac{1.0142}{2} \simeq 0.5071$$

and $-2\log BF_{01} \simeq 1.3581 < 2$; thus, there is no evidence in favor of H_0 nor in favour of H_1.

∎

Solution of Ex 2.
If

$$X_1, \ldots, X_n | P \overset{iid}{\sim} P \tag{1}$$
$$P | \alpha, P_0 \sim DP(\alpha, P_0), \tag{2}$$

then the marginal distribution of $(X_1, \ldots, X_n)$ can be found as

$$\begin{aligned}
\mathcal{L}(X_1, \ldots, X_n) &= \mathcal{L}(X_1) \times \mathcal{L}(X_2|X_1) \times \cdots \times \mathcal{L}(X_n|X_1, \ldots, X_{n-1}) \\
&= \prod_{i=1}^{n} \left(\frac{\alpha P_0(dX_i) + \sum_{j=1}^{i-1} \delta_{X_j}(dX_i)}{\alpha + i - 1}\right)
\end{aligned}$$

with the convention that the element in the brackets is P_0 when $i = 1$. This expression can be obtained by marginalizing out the random measure P from (**??**).

∎

BAYESIAN STATISTICS A. Guglielmi, I. Bianchini 29.01.2018

Properly justify all your answers.

Exercise 1 Answer this theoretical question first:

1. Let $Y_1, \dots, Y_m$ be i.i.d. random variables with distribution function $F_Y(y)$. Derive the distribution function of $X := Y_{(1)} = \min\{Y_1, \dots, Y_m\}$.

We aim at making inference on failure times of a device made of 5 components in a series system, so that the device (the system) *fails* if and only if anyone of the components *fails*. We assume that the system components, all of the same type A, work independently from each other, and that their failure times (in hours) $Y_1, \dots, Y_5$ have the exponential density with parameter $\theta > 0$, i.e.

$$f(y;\theta) = \theta e^{-\theta y} \mathbf{1}_{(0,\infty)}(y).$$

We bought 500 type A components and we built 100 series systems, monitoring the systems for 12 hours. At the end of this experiment, 13 systems were still working, and for the rest of the systems we got $\sum_{i=1}^{87} x_i = 867$ hours, where x_i denotes the lifetime of system i.

2. Derive the density of the random variable X representing the lifetime of a whole system.

3. Write the likelihood of all 100 observations. Then derive the conjugate prior to this likelihood, and find the hyperparameters of the posterior.

Censored data !!

4. Compute the prior hyperparameters such that the prior marginal mean and variance of the lifetime of each system are 5 and 100, respectively. Derive the corresponding posterior hyperparameters.

5. Compute the posterior predictive probability that a system will still be working after 12 hours.

6. Let[1] $\mu = E_\theta(X)$. Test the hypotheses:

$$H_0\colon \mu \geq 12 \text{ vs } H_1\colon \mu < 12,$$

computing the Bayes factor as the ratio between posterior and prior odds. Which hypothesis do you choose? Why?

(*Useful to know that d.f. of the gamma prior, evaluated at 1/60, is approx. 0.0575.*)

Exercise 2 Let $Z = |X|$, where X is a standard Gaussian r.v.

1. Derive the density $f_Z(z)$ of Z.

2. We build an acceptance rejection sampler to sample from $f_Z(z)$ with proposal density $g(z)$. Under what conditions does the sampler *work*?

3. Check the conditions you found in 2. when $g(z) = \lambda e^{-\lambda z} \mathbf{1}_{(0,+\infty)}$, where $\lambda > 0$.

4. Determine the acceptance probability and verify that its maximum is obtained when $\lambda = 1$.

5. Describe in detail the acceptance rejection sampler.

[1]Solve this point only if you finished the rest of the two exercises.

Solution of Ex 1.

1. It is straightforward to check that $X > x$ if and only if all r.v.s Y_i are larger than x. Hence:

$$F_X(x) = \mathbb{P}(X \le x) = 1 - \mathbb{P}(\min\{Y_1, \dots, Y_m\} > x) = 1 - \mathbb{P}(\{Y_1 > x\} \cap \dots \cap \{Y_m > x\})$$
$$= 1 - \prod_{i=1}^{m} \mathbb{P}(Y_i > x) = 1 - (1 - F_Y(x))^m, \tag{1}$$

since $Y_1 \dots, Y_m$ are i.i.d. with d.f. F_Y.

2. Recall that, if $Y \sim Exp(\theta)$, then $F_Y(y;\theta) = 1 - \mathrm{e}^{-\theta y}$ for $y \ge 0$, and $F_Y(y;\theta) = 0$ when $y < 0$. The system fails if and only if there is at least one component that fails. Hence the system lifetime X is given by the minimum of the component lifetimes. Using (??) we get:

$$F_X(x;\theta) = 1 - (\mathrm{e}^{-\theta x})^5 = 1 - \mathrm{e}^{-5\theta x}, \text{ if } x \ge 0 \text{ while } F_X(x;\theta) = 0, \text{ if } x < 0.$$

Hence $X \sim Exp(5\theta)$, so that $f_X(x;\theta) = 5\theta \mathrm{e}^{-5\theta x} \mathbf{1}_{(0,+\infty)}(x)$.

3. We also have censored data. Let $\boldsymbol{x} = (x_1, \dots, x_{87})$ denote observed lifetimes, and let $\boldsymbol{x}^* = (x_1^*, \dots, x_{13}^*)$ be censored data, i.e. those corresponding to non-observed failures; in this case we have $x_i^* = 12$ per $i = 1, \dots, 13$ and $\sum_1^{87} x_i = 867$. The likelihood, for $\theta > 0$, is

$$L(\theta; x_1, \dots, x_{87}, x_1^*, \dots, x_{13}^*) = \prod_{i=1}^{87} f_X(x_i;\theta) \prod_{i=1}^{13} (1 - F_x(x_i^*;\theta))$$
$$= \prod_{i=1}^{87} (5\theta \mathrm{e}^{-5\theta x_i}) \prod_{i=1}^{13} \mathrm{e}^{-5\theta x_i^*} = 5^{87} \theta^{87} \mathrm{e}^{-5\theta(\sum_{i=1}^{87} x_i + \sum_{i=1}^{13} x_i^*)}.$$

It is straightforward to see that this likelihood and the prior density $gamma(\alpha, \beta)$
$\pi(\theta) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \mathrm{e}^{-\beta\theta} \mathbf{1}_{(0,+\infty)}(\theta)$ are conjugate. Bayes's theorem gives the posterior density:

$$\pi(\theta; \underline{x}, \underline{x}^*) \propto \theta^{\alpha+87-1} \mathrm{e}^{-\theta(\beta + 5(\sum_{i=1}^{87} x_i + \sum_{i=1}^{13} x_i^*))}, \quad \theta > 0.$$

Hence posterior density for θ is $gamma(\alpha_1, \beta_1)$, where $\alpha_1 = \alpha + 87$ e $\beta_1 = \beta + 5(\sum_{i=1}^{87} x_i + \sum_{i=1}^{13} x_i^*))$.

4. For any fixed $i = 1, \dots, n$, X_i is the lifetime of system i. We have

$$5 = \mathrm{E}(X_i) = \mathrm{E}\left(E(X_i|\theta)\right) = \mathrm{E}(\frac{1}{5\theta}) = \frac{1}{5} \frac{\beta}{\alpha - 1} \Rightarrow \beta = 25(\alpha - 1)$$

and

$$100 = \mathrm{Var}(X_i) = \mathrm{Var}(\mathrm{E}(X_i|\theta)) + E(\mathrm{Var}(X_i|\theta)) = \mathrm{Var}\left[\frac{1}{5\theta}\right] + \mathrm{E}\left[\left(\frac{1}{5\theta}\right)^2\right]$$
$$= \frac{2}{25} \mathrm{Var}(\frac{1}{\theta}) + \frac{1}{25}\left(\mathrm{E}\left(\frac{1}{\theta}\right)^2\right) = \frac{2}{25} \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} + \frac{1}{25} \frac{\beta^2}{(\alpha-1)^2}$$
$$= \frac{1}{25} \frac{\beta^2}{(\alpha-1)^2} \left(\frac{2}{\alpha - 2} + 1\right) = \frac{1}{25} \frac{\alpha}{\alpha - 2} \frac{\beta^2}{(\alpha-1)^2} = 25 \frac{\alpha}{\alpha - 2}$$

so that

$$\frac{\alpha}{\alpha - 2} = 4 \Rightarrow \alpha = \frac{8}{3} \text{ and } \beta = \frac{125}{3}.$$

The parameters of the posterior distribution are equal to $\alpha_1 = \dfrac{8}{3} + 87 \simeq 89.76$ and $\beta_1 = \frac{125}{3} + 5115 \simeq 5156.67$.

5. Posterior predictive survival function is given by

$$1 - F_{X_{n+1}|X_1,\dots,X_n}(x;\underline{x},\underline{x}^*) := \mathbb{P}(X_{n+1} > x|\underline{x},\underline{x}^*) = \int_0^{+\infty} \mathbb{P}(X_{n+1} > x|\theta)\pi(\theta|\underline{x},\underline{x}^*)d\theta$$

$$= \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\int_0^{+\infty} \mathrm{e}^{-5\theta x}\theta^{\alpha_1-1}\mathrm{e}^{-\beta_1\theta}d\theta = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_1)}{(\beta_1+5x)^{\alpha_1}} = \left(\frac{\beta_1}{\beta_1+5x}\right)^{\alpha_1}.$$

   This yields that $1 - F_{X_{n+1}|X_1,\dots,X_n}(12;\underline{x},\underline{x}^*) = 0.354415$.

6. Since $\mu = \mathrm{E}_\theta(X) = (1/5\theta)$, H_0 is $\theta \le 1/60$ and $H_1 : \theta > 1/60$. We have to derive the Bayes factor between H_0 and H_1 as the ratio between posterior and prior odds:

$$BF_{01} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\mathbb{P}\left(\theta \le \frac{1}{60}|\underline{x},\underline{x}^*\right)/\mathbb{P}\left(\theta > \frac{1}{60}|\underline{x},\underline{x}^*\right)}{\mathbb{P}\left(\theta \le \frac{1}{60}\right)/\mathbb{P}\left(\theta > \frac{1}{60}\right)}.$$

   A priori $\theta \sim gamma(\alpha = 8/3, \beta = 125/3)$, so that the prior odds is

$$\mathbb{P}\left(\theta \le \frac{1}{60}\right)/\mathbb{P}\left(\theta > \frac{1}{60}\right) = 0.0575/(1-0.0575) \simeq 0.0610.$$

   As far as the posterior is concerned, we use CLT to approximate the df of the gamma distribution (with α large) by the Gaussian d.f. with same mean and variance. recall that

$$\mathrm{E}(\theta|\underline{x},\underline{x}^*) = \frac{\alpha_1}{\beta_1} = 0.0174 \quad \mathrm{Var}(\theta|\underline{x},\underline{x}^*) = \frac{\alpha_1}{\beta_1^2} = (0.0018)^2.$$

   Hence $\theta \sim gamma(\alpha_1, \beta_1) \simeq \mathcal{N}(0.0174, (0.0018)^2)$, so that the posterior odds is equal to:

$$\mathbb{P}\left(\theta \le \frac{1}{60}|\underline{x},\underline{x}^*\right)/\mathbb{P}\left(\theta > \frac{1}{60}|\underline{x},\underline{x}^*\right) \simeq \frac{\Phi\left(\frac{(1/60)-0.0174}{0.0018}\right)}{1-\Phi\left(\frac{(1/60)-0.0174}{0.0018}\right)} = \frac{0.3471}{1-0.3471} = 0.5317.$$

   In conclusion, $BF_{01} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{0.5317}{0.0610} = 8.7164$, e $2\log(BF_{01}) = 4.3304$. There is weak evidence in favour of H_0.

∎

Solution of Ex 2.

1. The distribution function of Z is 0 if $z \le 0$; when $z > 0$, then

$$P(Z \le z) = P(-z \le X \le z) = \Phi(z) - \Phi(-z),$$

   where $\Phi(\cdot)$ is the d.f. of the standard Gaussian r.v. X. Differencing we get:

$$f_Z(z) = \phi(z) + \phi(-z) = 2\phi(z) = \frac{2}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}, \text{ when } z > 0,$$

   and 0 otherwise, where $\phi(\cdot)$ is the density of the standard Gaussian r.v. X, which implies

$$f_Z(z) = \frac{2}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\mathbf{1}_{(0,+\infty)}(z)$$

2. There should exist a constant $M \geq 1$ such that

$$f_Z(z) \leq Mg(z) \text{ for all } z \in Supp(f_Z), \text{ and } Supp(g) \supseteq Supp(f_Z)$$

3. Let's start from the ratio between f_Z and g for $z > 0$:

$$\frac{f_Z(z)}{g(z)} = \frac{\frac{2}{\sqrt{2\pi}} e^{-z^2/2}}{\lambda e^{-\lambda z}} = \frac{2}{\lambda\sqrt{2\pi}} e^{\lambda z - z^2/2}$$

Its maximum can be derived solving the following inequality:

$$\frac{d}{dz}(\lambda z - z^2/2) = \lambda - z \geq 0 \Rightarrow z \leq \lambda,$$

which proves that the maximum of $(\lambda z - z^2/2)$, and consequently of $\frac{f_Z(z)}{g(z)}$, is achieved at $z = \lambda$. We have that, for all $z > 0$,

$$\frac{f_Z(z)}{g(z)} \leq \frac{2}{\lambda\sqrt{2\pi}} e^{\lambda^2 - \lambda^2/2} = \frac{2}{\lambda\sqrt{2\pi}} e^{\lambda^2/2} := M(\lambda)$$

Note that $Supp(f_Z) = Supp(g) = (0, +\infty)$.

4. The acceptance probability is

$$Q(\lambda) := \int_0^{+\infty} \frac{f_Z(z)}{Mg(z)} g(z)dz = \frac{1}{M(\lambda)} = \lambda\frac{\sqrt{2\pi}}{2} e^{-\lambda^2/2} < 1 \text{ since } M(\lambda) > 1$$

Its maximum is found solving the following inequality:

$$\frac{d}{dz}\left(\lambda\, e^{-\lambda^2/2}\right) = e^{-\lambda^2/2} + \lambda(-\lambda)e^{-\lambda^2/2} = e^{-\lambda^2/2}(1 - \lambda^2) \geq 0 \Rightarrow 0 \leq \lambda \leq 1.$$

Hence the maximum of the acceptance probability is obtained when $\lambda = 1$.

5. The acceptance-rejection algorithm for continuous random variables works as follows:

   (a) Generate a random variable Y distributed as g, namely sample from an exponential distribution with parameter $\lambda = 1$, using for instance the inverse transform sampling: generate a random variable V drawn from the uniform distribution on the unit interval (0, 1) and set $Y = -ln(V)$.

   (b) Generate U (independent from Y), sampling from a uniform distribution in $(0, 1)$.

   (c) If $U \leq \frac{f(Y)}{Mg(Y)}$, then set $Z = Y$ ("accept") ; otherwise go back to 1 ("reject"). When $\lambda = 1$ we have $M = e^{1/2}\frac{2}{\sqrt{2\pi}}$ and

$$\frac{f(Y)}{Mg(Y)} = e^{Y - Y^2/2 - 1/2}.$$

∎

BAYESIAN STATISTICS A. Guglielmi and I. Bianchini 13.09.2017

Properly justify all your answers.

Exercise 1 Consider the general case of an exponential family density

$$f(x|\theta) = h(x)\, e^{\theta^t T(x) - \Psi(\theta)},\ x \in \mathcal{X} \tag{1}$$

where θ is the *natural parameter*, $\theta \in \Theta \subset \mathbb{R}^s$, h is a positive function defined on $\mathcal{X}$, $T : \mathcal{X} \to \mathbb{R}^s$, $\Psi : \Theta \to \mathbb{R}^+$, and θ^t denotes the transpose (i.e. a row vector). The conjugate density for θ is

$$\pi(\theta; \mu, n_0) = K(\mu, n_0)\, e^{\theta^t \mu - n_0 \Psi(\theta)},\ \theta \in \Theta. \tag{2}$$

Consider a sample $(X_1, \ldots, X_N)$ from (**??**), i.e. the X_i's are independent and identically distributed random variables, given θ, according to (**??**).

1. Compute the posterior $\pi(\theta|x_1, \ldots, x_N)$. In addition, verify that prior (**??**) is conjugate to (**??**), and compute the hyperparameters of the posterior density, i.e. how μ and n_0 update in the posterior.

Henceforth, instead of the more general (**??**), we focus on the multinomial density, where, conditionally to the parameter, each X_i is a K-dimensional vector that sums up to n; its conditional density, for $\boldsymbol{x} := (x_1, x_2, \ldots, x_K)$, is

$$f(\boldsymbol{x}|(p_1, \ldots, p_K)) = \frac{n!}{x_1! \ldots x_K!} p_1^{x_1} \ldots p_K^{x_K},\ \text{if } x_i = 0, 1, \ldots, n \text{ for all } i,\ x_1 + \cdots + x_K = n, \tag{3}$$

and 0 otherwise; here $p_1, \ldots, p_K \in (0, 1)$, $p_K = 1 - (p_1 + \cdots + p_{K-1})$.

2. Show that (**??**) belongs to the exponential family, and write the corresponding expressions for the natural parameter $\boldsymbol{\theta}$, for $h(\cdot)$, $T(\cdot)$ and $\Psi(\cdot)$. Find the conjugate prior for $\boldsymbol{\theta}$ using (**??**).
3. Derive the conjugate prior for $(p_1, \ldots, p_{K-1}) \in S_{K-1} := \{p_1, \ldots, p_{K-1} \in (0,1), 0 < p_1 + \cdots + p_{K-1} < 1\}$ from point 3. Use the standard parameterization for prior hyperparameters.

   Consequently compute the posterior for $(p_1, \ldots, p_{K-1})$, given $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, the observed sample from (**??**). Derive the hyperparameters of the posterior $\pi(p_1, \ldots, p_{K-1}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$.

   Hint: *Use notation* $n_j := \sum_{i=1}^N x_{ij}$, $j = 1, \ldots, K$, *where* $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iK})^t$.
4. Compute the corresponding predictive density $f(\boldsymbol{x}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ for a new observation $\boldsymbol{X}_{N+1}$.

   Hint: *The following identity, valid for* $a_1, \ldots, a_k > 0$, *could be useful:*

$$\frac{\Gamma(a_1)\cdots\Gamma(a_{K-1})\Gamma(a_K)}{\Gamma(a_1 + \cdots + a_K)} = \int_{S_{K-1}} p_1^{a_1 - 1} \cdots p_{K-1}^{a_{K-1}-1}(1 - p_1 - \cdots p_{K-1})^{a_K - 1} dp_1 \cdots dp_{K-1}.$$

Assume henceforth $N = 1$. There are different species of tuna in the Mediterranean sea: common species tuna, yellowfin tuna and bluefin tuna. We assume density (**??**) for the counts of common species, yellowfin and bluefin tunas out of n subjects, under the conjugate prior derived at point 4.

5. Past data from biologists tell that, out of $m = 9$ tuna subjects, 4 were common tunas, 2 were yellowfin and 3 were bluefin tunas. Use the equivalent sample approach to fix the parameters of the prior found at point 4.
6. A fisherman go fishing in the sea and captures 6 tunas, equally shared among the three species. Compute the posterior, given the fisherman data, and write explicitly its hyperparameters.

7. Suppose that the fisherman will fish 6 tunas again. What is the predictive probability that the tunas will be equally shared among the three species, given available data?

Exercise 2 Let P be a Dirichlet process on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, i.e. $P \sim DP(a, P_0)$, where $a > 0$ and P_0 is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

1. Derive $\mathrm{E}(P(A))$ and $\mathrm{Var}(P(A))$ for every A in $\mathcal{B}(\mathbb{R})$. Moreover, for every A, B in $\mathcal{B}(\mathbb{R})$ with $A \cap B = \varnothing$, write the density of $(P(A), P(B))$, and compute $\mathrm{Cov}(P(A), P(B))$ with full detail.

2. Suppose now that P_0 is the standard Gaussian distribution. Fix a such that $\mathrm{Var}(P(-\infty, 0)) = 0.1$.

Solution of EX. 1

1. If $(x_1, \ldots, x_N)$ is the observed sample, we have

$$\begin{aligned}\pi(\theta|x_1, \ldots, x_N) &\propto \prod_1^N h(x_i) e^{\theta^t \sum_1^N T(x_i) - N\Psi(\theta)} \times e^{\theta^t \mu - n_0 \Psi(\theta)} \\ &\propto e^{\theta^t(\mu + \sum_1^N T(x_i)) - (n_0 + N)\Psi(\theta)}, \; \theta \in \Theta.\end{aligned}$$

It is clear from (??) that $\pi(\theta|x_1, \ldots, x_N) = \pi(\theta; \mu^{post}, n^{post})$, and

$$\mu^{post} = \mu + \sum_{i=1}^N T(x_i), \qquad n^{post} = n_0 + N;$$

this also means that the prior and the likelihood are conjugate.

2. The likelihood can be written in the exponential form as

$$p(\boldsymbol{x}|(p_1, \ldots, p_K)) = \frac{n!}{x_1! \ldots x_K!} \mathbf{1}_{Supp}(\boldsymbol{x}) e^{\log(p_1)x_1 + \cdots + \log(p_K)x_K}$$

where $Supp = \{x_i = 0, 1, \ldots, n \text{ for all } i, \; x_1 + \cdots + x_K = n\}$. Hence

$$\boldsymbol{\theta}^t = (\log(p_1), \ldots, \log(p_K)), \; T(\boldsymbol{x}) = (x_1, \ldots, x_K)^t, \; \Psi(\theta) = 0, \; h(\boldsymbol{x}) = \frac{n!}{x_1! \ldots x_K!} \mathbf{1}_{Supp}(\boldsymbol{x}).$$

The conjugate prior is thus given by

$$\pi(\boldsymbol{\theta}; \boldsymbol{\mu}) \propto e^{\boldsymbol{\mu}^t \boldsymbol{\theta}}.$$

3. From 2., the prior for $(p_1, \ldots, p_{K-1})$ is such that

$$\pi(\theta_1, \ldots, \theta_K | \mu) \propto p_1^{\mu_1} \ldots p_K^{\mu_K}, \; p_1, \ldots, p_K \in (0,1), p_K = 1 - (p_1 + \cdots + p_{K-1}),$$

which is a Dirichlet density on S_{K-1}.

We use the standard parameterization for the conjugate prior for $(p_1, \ldots, p_{K-1}) \in S_{K-1}$, that is $(p_1, \ldots, p_{K-1}) \sim Dir(a_1, \ldots, a_{K-1}, a_K)$, where $a_1, \ldots, a_K > 0$; its density is

$$\pi(p_1, \ldots, p_{K-1}; a_1, \ldots, a_K) = \frac{\Gamma(a_1 + \cdots + a_K)}{\Gamma(a_1) \ldots \Gamma(a_K)} p_1^{a_1 - 1} \cdots p_{K-1}^{a_{K-1} - 1} (1 - p_1 - \cdots p_{K-1})^{a_K - 1} \cdot \mathbf{1}_{S_{K-1}}(p_1, \ldots, p_{K-1}).$$

The corresponding posterior, given a sample $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N)$ of size N from (??), is

$$(p_1, \dots, p_{K-1})|\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \sim Dir(a_1^*, \dots, a_{K-1}^*, a_K^*) = Dir(a_1 + n_1, \dots, a_{K-1} + n_{K-1}, a_K + n_K),$$

where, for each $j = 1, \dots, K$, $n_j = \sum_{i=1}^N x_{ij}$ is the number of items of type j in the sample of size N (i.e. among $n \times N$ items), and $a_j^* = a_j + n_j$.

4. We have, for $\boldsymbol{x} = (x_1, \dots, x_K)$, $x_i = 0, 1, \dots, n$ for all i, $x_1 + \dots + x_K = n$:

$$\begin{aligned} p(\boldsymbol{x}|x_1, \dots, x_N) &= \int_{S_{K-1}} f(\boldsymbol{x}|p_1, \dots, p_{K-1})\pi(p_1, \dots, p_{K-1}|\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) dp_1 \cdots dp_{K-1} \\ &= \int_{S_{K-1}} \frac{n!}{x_1! \dots x_K!} p_1^{x_1} \dots p_{K-1}^{x_{K-1}} (1 - p_1 - \dots - p_{K-1})^{x_K} \\ &\quad \times \frac{\Gamma(a_1^* + \dots + a_K^*)}{\Gamma(a_1^*) \dots \Gamma(a_K^*)} p_1^{a_1^*-1} \cdots p_{K-1}^{a_{K-1}^*-1} (1 - p_1 - \cdots p_{K-1})^{a_K^*-1} dp_1 \cdots dp_{K-1} \\ &= \frac{n!}{x_1! \dots x_K!} \frac{\Gamma(a_1^* + \dots + a_K^*)}{\Gamma(a_1^*) \dots \Gamma(a_K^*)} \frac{\Gamma(a_1^* + x_1) \cdots \Gamma(a_K^* + x_K)}{\Gamma(a_1^* + x_1 + \dots + a_K^* + x_K)} \\ &= \frac{n!}{x_1! \dots x_K!} \frac{\Gamma(a_1^* + \dots + a_K^*)}{\Gamma(a_1^*) \dots \Gamma(a_K^*)} \frac{\Gamma(a_1^* + x_1) \cdots \Gamma(a_K^* + x_K)}{\Gamma(a_1^* + \dots + a_K^* + n)} \end{aligned}$$

5. Now $N = 1$ and $K = 3$; using past information ($m = 9$, $m_1 = 4$, $m_2 = 2$, $m_3 = 3$), we have

$$\mathbb{E}(p_j|\text{past data}) = \frac{a_j + m_j}{\alpha + m} = \frac{a_j}{\alpha} \frac{\alpha}{\alpha + m} + \frac{m_j}{m} \frac{m}{\alpha + m}, \; j = 1, 2, 3$$

where $\alpha = \sum_{l=1}^3 a_l$. Here p_1, p_2, p_3 are the probabilities that a tuna is from the common, yellowfin, and bluefin species, respectively.

It is easy to conclude that $a_j \leftrightarrow m_j$ and $\alpha \leftrightarrow m$: therefore, $(a_1, a_2, a_3) = (4, 2, 3)$.

6. The data from the fisherman amount to $n = 6$, $n_1 = 2$, $n_2 = 2$ $n_3 = 2$. Hence the posterior of (p_1, p_2) is still Dirichlet with parameters $(4 + 2, 2 + 2, 3 + 2) = (6, 4, 5) = (a_1^*, a_2^*, a_3^*)$.

7. From 4., the predictive distribution is

$$p(\text{"2 tunas of each of the 3 species out of 6"}|data) = \frac{6!}{2!2!2!} \frac{\Gamma(a_1^* + a_2^* + a_3^*)}{\Gamma(a_1^*)\Gamma(a_2^*)\Gamma(a_3^*)} \frac{\Gamma(a_1^* + 2)\Gamma(a_2^* + 2)\Gamma(a_3^* + 2)}{\Gamma(a_1^* + a_2^* + a_3^* + 6)}$$

$$= \frac{6!}{2!2!2!} \frac{\Gamma(15)}{\Gamma(6)\Gamma(4)\Gamma(5)} \frac{\Gamma(8)\Gamma(6)\Gamma(7)}{\Gamma(21)} = \frac{6!}{2!2!2!} \frac{14! \times 7 \times 6 \times 5 \times 4 \times 6 \times 5}{20!} = \frac{105}{1292} \simeq 0.081269.$$

∎

Solution of EX. 2 For the properties of the Dirichlet process, we know that, for every $A \in \mathcal{B}(\mathbb{R})$, $(P(A), P(A^c)) \sim Dirichlet(aP_0(A), aP_0(A^c))$, so that

$$P(A) \sim Beta(aP_0(A), a(1 - P_0(A)))$$

since $P(A^c) = 1 - P(A)$.

1. Therefore,

$$\mathbb{E}(P(A)) = \frac{aP_0(A)}{aP_0(A) + a(1 - P_0(A))} = P_0(A)$$

and

$$
\begin{aligned}
\mathrm{Var}(P(A)) &= \frac{aP_0(A) \times a(1-P_0(A))}{(aP_0(A)+a(1-P_0(A)))^2 \times (aP_0(A)+a(1-P_0(A))+1)} \\
&= \frac{a^2 P_0(A)(1-P_0(A))}{a^2(a+1)} = \frac{P_0(A)(1-P_0(A))}{a+1}.
\end{aligned}
$$

Consider now $A, B \in \mathcal{B}(\mathbb{R})$, $A \cap B = \varnothing$: the partition generated by A and B is $A, B, A^c \cap B^c$, and, by the definition of the Dirichlet process,

$$
(P(A), P(B), P(A^c \cap B^c)) \sim \mathit{Dirichlet}(aP_0(A), aP_0(B), aP_0(A^c \cap B^c)).
$$

Then, $\mathrm{Cov}(P(A), P(B)) = \mathbb{E}(P(A)P(B)) - P_0(A)P_0(B)$. We have:

$$
\begin{aligned}
\mathbb{E}(P(A)P(B)) &= \int_D xy \frac{\Gamma(a) x^{aP_0(A)-1} y^{aP_0(B)-1}}{\Gamma(aP_0(A))\Gamma(aP_0(B))\Gamma(aP_0(A^c \cap B^c))} (1-x-y)^{aP_0(A^c \cap B^c)-1} dxdy \\
&= \frac{\Gamma(a)}{\Gamma(aP_0(A))\Gamma(aP_0(B))\Gamma(aP_0(A^c \cap B^c))} \frac{\Gamma(aP_0(A)+1)\Gamma(aP_0(B)+1)\Gamma(aP_0(A^c \cap B^c))}{\Gamma(a+2)} \\
&= \frac{aP_0(A)\; aP_0(B)}{a(a+1)} = \frac{a}{a+1} P_0(A)\; P_0(B).
\end{aligned}
$$

where $D = \{x, y > 0 : 0 < x + y < 1\}$. Consequently:

$$
\mathrm{Cov}(P(A), P(B)) = P_0(A)P_0(B)\left(\frac{a}{a+1} - 1\right) = -\frac{P_0(A)P_0(B)}{a+1}.
$$

2. From 1.,

$$
0.1 = \mathrm{Var}(P(A)) = \frac{P_0(A)(1-P_0(A))}{a+1} = \frac{\frac{1}{4}}{a+1} = \frac{1}{4(a+1)}
$$

so that

$$
4(a+1) = 10 \Rightarrow a+1 = \frac{5}{2} \Rightarrow a = \frac{3}{2}.
$$

∎

BAYESIAN STATISTICS A. Guglielmi and I. Bianchini 19.07.2017

Properly justify all your answers.

Exercise 1 Consider the general case of an exponential family density

$$f(x|\theta) = h(x)\,\mathrm{e}^{\theta^t T(x) - \Psi(\theta)},\ x \in \mathcal{X} \tag{1}$$

where θ is the *natural parameter*, $\theta \in \Theta \subset \mathbb{R}^s$, h is a positive function defined on $\mathcal{X}$, $T : \mathcal{X} \to \mathbb{R}^s$, $\Psi : \Theta \to \mathbb{R}^+$, and θ^t denotes the transpose (i.e. a row vector). The conjugate density for θ is

$$\pi(\theta; \mu, n_0) = K(\mu, n_0)\,\mathrm{e}^{\theta^t \mu - n_0 \Psi(\theta)},\ \theta \in \Theta. \tag{2}$$

Consider a sample $(X_1, \ldots, X_N)$ from (1), i.e. the X_i's are independent and identically distributed random variables, given θ, according to (1).

1. Compute the posterior $\pi(\theta|x_1, \ldots, x_N)$. In addition, verify that prior (2) is conjugate to (1), and compute the hyperparameters of the posterior density, i.e. how μ and n_0 update in the posterior.

Henceforth, instead of the more general (1), we focus on the multinomial density, where, conditionally to the parameter, each X_i is a K-dimensional vector that sums up to n; its conditional density, for $\boldsymbol{x} := (x_1, x_2, \ldots, x_K)$, is

$$f(\boldsymbol{x}|(p_1, \ldots, p_K)) = \frac{n!}{x_1! \ldots x_K!} p_1^{x_1} \ldots p_K^{x_K},\ \text{if } x_i = 0, 1, \ldots, n \text{ for all } i,\ x_1 + \cdots + x_K = n, \tag{3}$$

and 0 otherwise; here $p_1, \ldots, p_K \in (0, 1)$, $p_K = 1 - (p_1 + \cdots + p_{K-1})$.

2. Show that (3) belongs to the exponential family, and write the corresponding expressions for the natural parameter $\boldsymbol{\theta}$, for $h(\cdot)$, $T(\cdot)$ and $\Psi(\cdot)$. Find the conjugate prior for $\boldsymbol{\theta}$ using (2).

3. Derive the conjugate prior for $(p_1, \ldots, p_{K-1}) \in S_{K-1} := \{p_1, \ldots, p_{K-1} \in (0, 1), 0 < p_1 + \cdots + p_{K-1} < 1\}$ from point 3. Use the standard parameterization for prior hyperparameters.

   Consequently compute the posterior for $(p_1, \ldots, p_{K-1})$, given $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, the observed sample from (3). Derive the hyperparameters of the posterior $\pi(p_1, \ldots, p_{K-1}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$.

   Hint: *Use notation* $n_j := \sum_{i=1}^N x_{ij}$, $j = 1, \ldots, K$, *where* $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iK})^t$.

4. Compute the corresponding predictive density $f(\boldsymbol{x}|\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ for a new observation $\boldsymbol{X}_{N+1}$.

   Hint: *The following identity, valid for* $a_1, \ldots, a_k > 0$, *could be useful:*

$$\frac{\Gamma(a_1) \cdots \Gamma(a_{K-1})\Gamma(a_K)}{\Gamma(a_1 + \cdots + a_K)} = \int_{S_{K-1}} p_1^{a_1 - 1} \cdots p_{K-1}^{a_{K-1}-1}(1 - p_1 - \cdots p_{K-1})^{a_K - 1} dp_1 \cdots dp_{K-1}.$$

Assume henceforth $N = 1$. There are different species of tuna in the Mediterranean sea: common species tuna, yellowfin tuna and bluefin tuna. We assume density (3) for the counts of common species, yellowfin and bluefin tunas out of n subjects, under the conjugate prior derived at point 4.

5. Past data from biologists tell that, out of $m = 10$ tuna subjects, 5 were common tunas, 3 were yellowfin and 2 were bluefin tunas. Use the equivalent sample approach to fix the parameters of the prior found at point 4.

6. A fisherman go fishing in the sea and captures 5 tunas: 2 of them are from the common species, 1 is yellowfin and 2 are bluefin. Compute the posterior, given the fisherman data, and write explicitly its hyperparameters.

7. Suppose that the fisherman will fish 5 tunas again. What is the predictive probability that the tunas will all be of bluefin type, given available data?

Exercise 2 Give the definition of conditional predictive ordinate (CPO) for each observation y_i in a sample $y = (y_1, \ldots, y_n)$, where $Y_1, \ldots, Y_n$ are independent conditionally to the parameter θ. In particular, denote by $\boldsymbol{y}^{-i}$ the vector containing all the observations but the i-th, and by $f_i(y_i; \theta)$, $\theta \in \Theta$, the density of the conditional distribution of the i-th observation, given θ.
Describe how to use the values $CPO_1, \ldots, CPO_n$ to define an index for assessing the predictive goodness-of-fit of the model.
Finally, suppose to have a sample from a MCMC sample $\{\theta^{(g)}, g = 1, \ldots, G\}$ from the posterior of θ given $(y_1, \ldots, y_n)$; describe how can you use it to estimate CPO_i for each $i = 1, \ldots, n$.

Solution of EX. 1

1. If $(x_1, \ldots, x_N)$ is the observed sample, we have

$$\pi(\theta|x_1, \ldots, x_N) \propto \prod_1^N h(x_i) e^{\theta^t \sum_1^N T(x_i) - N\Psi(\theta)} \times e^{\theta^t \mu - n_0 \Psi(\theta)}$$
$$\propto e^{\theta^t(\mu + \sum_1^N T(x_i)) - (n_0+N)\Psi(\theta)}, \ \theta \in \Theta.$$

   It is clear from (2) that $\pi(\theta|x_1, \ldots, x_N) = \pi(\theta; \mu^{post}, n^{post})$, and

$$\mu^{post} = \mu + \sum_{i=1}^N T(x_i), \quad n^{post} = n_0 + N;$$

   this also means that the prior and the likelihood are conjugate.

2. The likelihood can be written in the exponential form as

$$p(\boldsymbol{x}|(p_1, \ldots, p_K)) = \frac{n!}{x_1! \ldots x_K!} \mathbf{1}_{Supp}(\boldsymbol{x}) e^{\log(p_1)x_1 + \cdots + \log(p_K)x_K}$$

   where $Supp = \{x_i = 0, 1, \ldots, n \text{ for all } i, \ x_1 + \cdots + x_K = n\}$. Hence

$$\boldsymbol{\theta}^t = (\log(p_1), \ldots, \log(p_K)), \ T(\boldsymbol{x}) = (x_1, \ldots, x_K)^t, \ \Psi(\theta) = 0, \ h(\boldsymbol{x}) = \frac{n!}{x_1! \ldots x_K!} \mathbf{1}_{Supp}(\boldsymbol{x}).$$

   The conjugate prior is thus given by

$$\pi(\boldsymbol{\theta}; \boldsymbol{\mu}) \propto e^{\boldsymbol{\mu}^t \boldsymbol{\theta}}.$$

3. From 2., the prior for $(p_1, \ldots, p_{K-1})$ is such that

$$\pi(\theta_1, \ldots, \theta_K|\mu) \propto p_1^{\mu_1} \ldots p_K^{\mu_K}, \ p_1, \ldots, p_K \in (0,1), p_K = 1 - (p_1 + \cdots + p_{K-1}),$$

   which is a Dirichlet density on S_{K-1}.

4. We use the standard parameterization for the conjugate prior for $(p_1, \ldots, p_{K-1}) \in S_{K-1}$, that is $(p_1, \ldots, p_{K-1}) \sim Dir(a_1, \ldots, a_{K-1}, a_K)$, where $a_1, \ldots, a_K > 0$; its density is

$$\pi(p_1, \ldots, p_{K-1}; a_1, \ldots, a_K) = \frac{\Gamma(a_1 + \cdots + a_K)}{\Gamma(a_1) \ldots \Gamma(a_K)} p_1^{a_1 - 1} \cdots p_{K-1}^{a_{K-1}-1} (1 - p_1 - \cdots p_{K-1})^{a_K - 1} \cdot \mathbf{1}_{S_{K-1}}(p_1, \ldots, p_{K-1}).$$

The corresponding posterior, given a sample $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ of size N from (3), is

$$(p_1, \ldots, p_{K-1}) | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \sim Dir(a_1^*, \ldots, a_{K-1}^*, a_K^*) = Dir(a_1 + n_1, \ldots, a_{K-1} + n_{K-1}, a_K + n_K),$$

where, for each $j = 1, \ldots, K$, $n_j = \sum_{i=1}^N x_{ij}$ is the number of items of type j in the sample of size N (i.e. among $n \times N$ items), and $a_j^* = a_j + n_j$.

5. We have, for $\boldsymbol{x} = (x_1, \ldots, x_K)$, $x_i = 0, 1, \ldots, n$ for all i, $x_1 + \cdots + x_K = n$:

$$p(\boldsymbol{x} | x_1, \ldots, x_N) = \int_{S_{K-1}} f(\boldsymbol{x} | p_1, \ldots, p_{K-1}) \pi(p_1, \ldots, p_{K-1} | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) dp_1 \cdots dp_{K-1}$$

$$= \int_{S_{K-1}} \frac{n!}{x_1! \ldots x_K!} p_1^{x_1} \ldots p_{K-1}^{x_{K-1}} (1 - p_1 - \cdots - p_{K-1})^{x_K}$$

$$\times \frac{\Gamma(a_1^* + \cdots + a_K^*)}{\Gamma(a_1^*) \ldots \Gamma(a_K^*)} p_1^{a_1^* - 1} \cdots p_{K-1}^{a_{K-1}^* - 1} (1 - p_1 - \cdots p_{K-1})^{a_K^* - 1} dp_1 \cdots dp_{K-1}$$

$$= \frac{n!}{x_1! \ldots x_K!} \frac{\Gamma(a_1^* + \cdots + a_K^*)}{\Gamma(a_1^*) \ldots \Gamma(a_K^*)} \frac{\Gamma(a_1^* + x_1) \cdots \Gamma(a_K^* + x_K)}{\Gamma(a_1^* + x_1 + \cdots + a_K^* + x_K)}$$

$$= \frac{n!}{x_1! \ldots x_K!} \frac{\Gamma(a_1^* + \cdots + a_K^*)}{\Gamma(a_1^*) \ldots \Gamma(a_K^*)} \frac{\Gamma(a_1^* + x_1) \cdots \Gamma(a_K^* + x_K)}{\Gamma(a_1^* + \cdots + a_K^* + n)}$$

6. Now $N = 1$ and $K = 2$; using past information ($m = 10$, $m_1 = 5$, $m_2 = 3$, $m_3 = 2$), we have

$$\mathbb{E}(p_j | \text{past data}) = \frac{a_j + m_j}{\alpha + m} = \frac{a_j}{\alpha} \frac{\alpha}{\alpha + m} + \frac{m_j}{m} \frac{m}{\alpha + m}, \; j = 1, 2, 3$$

where $\alpha = \sum_{l=1}^3 a_l$. Here p_1, p_2, p_3 are the probabilities that a tuna is from the common, yellowfin, and bluefin species, respectively.

It is easy to conclude that $a_j \leftrightarrow m_j$ and $\alpha \leftrightarrow m$: therefore, $(a_1, a_2, a_3) = (5, 3, 2)$.

7. The data from the fisherman amount to $n = 5$, $n_1 = 2$, $n_2 = 1$ $n_2 = 2$. Hence the posterior of (p_1, p_2) is still Dirichlet with parameters $(5 + 2, 3 + 1, 2 + 2) = (7, 4, 4) = (a_1^*, a_2^*, a_3^*)$.

8. From 5., the predictive distribution is

$$p(\text{“5 bluefin tunas out of 5”} | data) = \frac{5!}{0!0!5!} \frac{\Gamma(a_1^* + a_2^* + a_3^*)}{\Gamma(a_1^*)\Gamma(a_2^*)\Gamma(a_3^*)} \frac{\Gamma(a_1^*)\Gamma(a_2^*)\Gamma(a_3^* + 5)}{\Gamma(a_1^* + a_2^* + a_3^* + 5)}$$

$$= \frac{\Gamma(a_1^* + a_2^* + a_3^*)}{\Gamma(a_1^* + a_2^* + a_3^* + 5)} \frac{\Gamma(a_3^* + 5)}{\Gamma(a_3^*)} = \frac{\Gamma(15)}{\Gamma(20)} \frac{\Gamma(9)}{\Gamma(4)} = \frac{14!}{19!} \frac{8!}{3!} \simeq 0.004815961.$$

∎

Solution of EX. 2 See Section 4.9.2 of the book of Christensen et al, *Bayesian ideas and data analysis*, 2011.

∎

BAYESIAN STATISTICS A. Guglielmi and I. Bianchini 05.07.2017

Properly justify all your answers.

Exercise 1 Let $X_1, X_2, \ldots, X_n$ be (conditionally) independent and identically distributed random variables following the inverse-Gaussian density with parameters $\mu, \tau > 0$:

$$f(x;\mu,\tau) = \sqrt{\frac{\tau}{2\pi x^3}} \exp\left\{-\frac{\tau\mu^2}{2x}(x-\frac{1}{\mu})^2\right\} \mathbb{I}_{(0,+\infty)}(x). \tag{1}$$

Note that, for any $i = 1, \ldots, n$,

$$\mathrm{E}(X_i|\mu,\tau) = \frac{1}{\mu}, \quad \mathrm{Var}(X_i|\mu,\tau) = \frac{1}{\tau\mu^3}.$$

1. Compute the likelihood $L(\mu,\tau;x_1,\ldots,x_n)$ for non-negative datapoints $(x_1,\ldots,x_n)$.
2. For (μ,τ), assume the prior density:

$$\begin{aligned}\pi(\mu,\tau|\nu,\alpha=1,\beta=\frac{1}{\nu}) &= \kappa(\nu,\alpha,\beta)\tau^{\nu/2-1}\exp\left\{-\frac{\nu\alpha\tau}{2}\left[1+\frac{\beta}{\alpha}\left(\mu-\frac{1}{\beta}\right)^2\right]\right\}\mathbf{1}_{(0,+\infty)^2}(\mu,\tau)\\ &= \widetilde{\kappa}(\nu)\tau^{\nu/2-1}\exp\left\{-\frac{\nu\tau}{2}\left[1+\frac{1}{\nu}(\mu-\nu)^2\right]\right\}\mathbf{1}_{(0,+\infty)^2}(\mu,\tau) \end{aligned} \tag{2}$$

   where $\nu > 1$, and $\kappa(\nu,\alpha,\beta)$, $\widetilde{\kappa}(\nu)$ are normalizing constants. Verify that (2) is conjugate to the likelihood found at point 1., and compute explicitly the posterior density. In particular, derive the hyperparameters ν_n, α_n, β_n of the posterior $\pi(\mu,\tau|\nu_n,\alpha_n,\beta_n)$.

Assume henceforth that μ **is known** and **equal to 1**.

3. Write down the explicit expression of the prior density

$$\pi_1(\tau|\nu) := \pi(1,\tau|\nu,\alpha=1,\beta=\frac{1}{\nu}) \text{ as in (2)}$$

   and the corresponding posterior density $\pi_1(\tau|data) := \pi(\mu=1,\tau|\nu_n,\alpha_n,\beta_n)$. Are these densities of known form? Which are the hyperparameters a priori and a posteriori?
4. Fix ν in $\pi_1(\tau|\nu)$ henceforth so that a priori (marginally) $\mathrm{Var}(X_1) = 11.375$. If there are many such ν's, choose the largest value.

For failure times $X_1,\ldots,X_{100}$ in the context of reliability of certain electronic devices under high stress condition, we assume the conditional distribution (1) (with $\mu = 1$); summary statistics of the data are the following:

$$\sum_1^{100}\frac{1}{x_i} = 857.5 \quad \sum_1^{100} x_i = 97\ .$$

5. Under the available data, compute the hyperparameters of the posterior density $\pi_1(\tau|data)$ derived at point 3. with ν as determined at point 4.
6. By using the *Bayes factor*, test hypotheses $H_0 : \tau \le \tau_0 := 0.17$ vs $H_1 : \tau > \tau_0$ when the prior is the density π_1 found at points 3. and 4. Which decision does the Bayes factor support?

   Hint: use the Gaussian density to approximate the posterior, if appropriate. Moreover, it would be useful to know that $\pi_1(\tau \le 0.17|\nu) = 0.8842$ with ν as determined at point 4.

Exercise 2 Consider the probit regression model

$$\begin{aligned} Y_i|\boldsymbol{x}_i, \boldsymbol{\beta} &\overset{\text{ind.}}{\sim} Be(p_i), \ i = 1, \dots, n \\ p_i &= \Phi(\boldsymbol{x}_i^t \boldsymbol{\beta}) = \Phi(\beta_1 x_{i1} + \dots + \beta_p x_{ip}) \\ \boldsymbol{\beta} &\sim \mathcal{N}_p(\boldsymbol{b}_0, B_0), \end{aligned} \tag{3}$$

where $\boldsymbol{b}_0 \in \mathbb{R}^p$, B_0 is a $p \times p$ covariance matrix and Φ is the standard gaussian d.f.. Introducing suitable latent variables, describe a Gibbs sampler algorithm to simulate from the posterior distribution of $\boldsymbol{\beta}$, given data $y_1, \dots, y_n$, according to model (3).

Solution of EX. 1

1. For $x_1, \dots, x_n > 0$, the likelihood is

$$\mathcal{L}(\mu, \tau; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \mu, \tau) \propto \tau^{n/2} \exp\{-\frac{\tau}{2}(\mu^2 \sum_1^n x_i + \sum_1^n \frac{1}{x_i} - 2n\mu)\} \mathbf{1}_{(0,+\infty)^2}(\mu, \tau).$$

2. The posterior density is proportional to

$$\pi(\mu, \tau | x_1, \dots, x_n) \propto \tau^{n/2} \exp\{-\frac{\tau}{2}(\mu^2 \sum_1^n x_i + \sum_1^n \frac{1}{x_i} - 2n\mu)\} \times \tau^{\nu/2-1} \exp\left\{-\frac{\nu\tau}{2}\left[1 + \frac{1}{\nu}(\mu - \nu)^2\right]\right\} \times \mathbf{1}_{(0,+\infty)^2}(\mu, \tau)$$

$$\propto \tau^{\frac{\nu+n}{2}-1} \exp\left\{-\frac{\tau}{2}(1 + \sum x_i)\left(\mu - \frac{\nu+n}{1+\sum x_i}\right)^2\right\}$$
$$\times \exp\left\{-\frac{\tau}{2}(1 + \sum x_i)\left(\frac{\sum \frac{1}{x_i} + \nu^2 + \nu}{1 + \sum x_i} - \left(\frac{\nu+n}{1+\sum x_i}\right)^2\right)\right\} \times \mathbf{1}_{(0,+\infty)^2}(\mu, \tau).$$

Hence, if we assume the following notation:

$$\nu_n = \nu + n, \ \beta_n = \frac{1 + \sum x_i}{\nu + n},$$
$$\alpha_n = \beta_n \left(\frac{\sum \frac{1}{x_i} + \nu^2 + \nu}{1 + \sum x_i} - \left(\frac{\nu+n}{1+\sum x_i}\right)^2\right) = \frac{1 + \sum x_i}{\nu + n}\left(\frac{\sum \frac{1}{x_i} + \nu^2 + \nu}{1 + \sum x_i} - \left(\frac{\nu+n}{1+\sum x_i}\right)^2\right)$$
$$= \frac{\sum \frac{1}{x_i} + \nu^2 + \nu}{\nu_n} - \frac{1}{\beta_n},$$

we have that

$$\pi(\mu, \tau | x_1, \dots, x_n) \propto \tau^{\nu_n/2-1} \exp\left\{-\frac{\nu_n \alpha_n \tau}{2}\left[1 + \frac{\beta_n}{\alpha_n}\left(\mu - \frac{1}{\beta_n}\right)^2\right]\right\} \mathbf{1}_{(0,+\infty)^2}(\mu, \tau),$$

i.e. (2) is conjugate.

3. From the posterior (2), substituting $\mu = 1$ in that expression, we have:

$$\pi_1(\tau|\nu) = \pi(1, \tau | \nu, \alpha = 1, \beta = \frac{1}{\nu}) \propto \tau^{\nu/2-1} \exp\left\{-\frac{\nu\tau}{2}\left[1 + \frac{1}{\nu}(1 - \nu)^2\right]\right\} \mathbf{1}_{(0,+\infty)^2}(\mu, \tau)$$
$$\propto \tau^{\nu/2-1} \exp\left\{-\frac{\tau}{2}\left[\nu + (1 - \nu)^2\right]\right\} \mathbf{1}_{(0,+\infty)^2}(\mu, \tau),$$

that is a gamma density with hyperparameters $a = \frac{\nu}{2}$, $b = \frac{\nu}{2} + \frac{(\nu-1)^2}{2}$. The posterior density is:

$$\pi(\mu, \tau | x_1, \ldots, x_n) \propto \tau^{n/2} \exp\{-\frac{\tau}{2}(\sum_1^n x_i + \sum_1^n \frac{1}{x_i} - 2n)\} \times \tau^{\nu/2-1} \exp\left\{-\frac{\nu\tau}{2}\left[1 + \frac{1}{\nu}(1-\nu)^2\right]\right\}$$

From the expression of the posterior derived at point 2. (with $\mu = 1$), we got that the corresponding posterior is a gamma density with hyperparameters

$$a_n = a + \frac{n}{2} = \frac{\nu + n}{2},\ b_n = b + \frac{1}{2}(\sum_1^n x_i + \sum_1^n \frac{1}{x_i} - 2n) = \frac{\nu + (\nu-1)^2 + \sum_1^n x_i + \sum_1^n \frac{1}{x_i} - 2n}{2}.$$

4. Since $\mathrm{E}(X_i) = \mathrm{E}\left(E(X_i|\tau)\right) = 1$, and, for $\nu > 2$, $\mathrm{Var}(X_i) = \mathrm{Var}(\mathrm{E}(X_i|\mu,\tau)) + E(\mathrm{Var}(X_i|\mu,\tau)) = \mathrm{E}\left(\frac{1}{\tau}\right) = \frac{\beta}{\alpha - 1} = \frac{\nu + (\nu-1)^2}{\nu - 2}$, we set ν so that

$$\frac{\nu + (\nu-1)^2}{\nu-2} = 11.375 \Rightarrow \nu^2 - 12.375\nu + 23.75 = 0.$$

There are two solutions of this equation (both satisfying $\nu > 2$), i.e. 2.375 and 10, so that we choose $\nu = 10$. In this case, the prior $\pi_1(\tau|\nu)$ is the gamma density with parameter $a = \frac{\nu}{2} = 5$, $b = \frac{\nu}{2} + \frac{(\nu-1)^2}{2} = 45.5$.

5. The posterior is the gamma density with hyperparameters

$$a_n = a + \frac{n}{2} = 5 + \frac{100}{2} = 55,\ b_n = b + \frac{1}{2}(\sum_1^n x_i + \sum_1^n \frac{1}{x_i} - 2n) = 422.75.$$

6. Verify $H_0 : \tau \leq \tau_0 = 0.17$ vs $H_1 : \tau > \tau_0$, where a posteriori

$$\tau | data \sim gamma(\alpha_n = 55, \beta_n = 422.75) \simeq \mathcal{N}(\frac{\alpha_n}{\beta_n} = 0.1301, \frac{\alpha_n}{\beta_n^2} = (0.0175)^2).$$

The Bayes factor between H_0 and H_1 is

$$BF_{01} = \frac{\frac{\pi_1(H_0|\mathbf{x})}{\pi_1(H_1|\mathbf{x})}}{\frac{\pi_1(H_0)}{\pi_1(H_1)}}.$$

We computed the numerator of the previous fraction using the Gaussian approximation $\mathcal{N}\left(\frac{\alpha_n}{\beta_n}, \frac{\alpha_n}{\beta_n^2}\right) = \mathcal{N}(0.1301, (0.0175)^2)$:

$$\text{gamma df}(0.17, \alpha_n = 55, \beta_n = 422.7537) \simeq \Phi(\frac{0.17 - 0.1301}{0.0175}) \simeq 0.9886962 \simeq 0.9887;$$

the prior odds is given by $\frac{0.8842}{1 - 0.8842} \simeq 7.6324$, hence

$$BF_{01} \simeq 11.45893 \Rightarrow 2\log(BF_{01}) \simeq 4.877539\ .$$

The exact value of the posterior odds is

$$\frac{0.9830}{1 - 0.9830} = 57.82353 \Rightarrow 2\log(BF_{01}) = 4.051043.$$

In any case, there is evidence in favour of H_0.

Solution of EX. 2 We introduce latent variables $Z_1, \dots, Z_n$ such that

$$Y_i|\boldsymbol{x}_i = \begin{cases} 1 & \text{if } Z_i > 0 \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \dots, n \tag{4}$$

$$Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{x}_i^t \boldsymbol{\beta}, 1). \tag{5}$$

It is straightforward to check that the likelihood in (3) is equivalent to (4)-(5). The *parameter* is now $(\boldsymbol{\beta}, \boldsymbol{Z})$, where $\boldsymbol{Z} = (Z_1, \dots, Z_n)^t$. We build a Gibbs sampler to simulate from the joint posterior law $\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{Z}|y_1, \dots, y_n)$, but in the end we will be interested only in the marginal posterior $\mathcal{L}(\boldsymbol{\beta}|y_1, \dots, y_n)$. In order to derive a Gibbs sampler, we need the full-conditionals, that are proportional to the joint law of data and the parameter vector:

$$\mathcal{L}(Y_1, \dots, Y_n, \boldsymbol{\beta}, \boldsymbol{Z}) \propto \mathcal{L}(Y_1, \dots, Y_n|\boldsymbol{\beta}, \boldsymbol{Z}) \times \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{Z}) = \mathcal{L}(Y_1, \dots, Y_n|\boldsymbol{Z}) \times \mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}) \times \pi(\boldsymbol{\beta}).$$

Then, the full-conditionals are:

- $\mathcal{L}(\boldsymbol{\beta}|\boldsymbol{Z}, \boldsymbol{Y}) \propto \mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}) \times \pi(\boldsymbol{\beta})$. This last distribution is the posterior of a linear model with Gaussian likelihood, where the *data* are $Z_1, \dots, Z_n$, i.e.

  $$\mathcal{L}(\boldsymbol{\beta}|\boldsymbol{Z}, \boldsymbol{Y}) = \mathcal{N}_p(\boldsymbol{b}_n, B_n), \quad B_n = (X^t X + B_0^{-1})^{-1}, \ \boldsymbol{b}_n = B_n(X^t X \hat{\boldsymbol{\beta}} + B_0^{-1} \boldsymbol{b}_0).$$

  Here X is the design matrix with $\boldsymbol{x}_i$'s as rows, and $\hat{\boldsymbol{\beta}} = (X^t X)^{-1} X^t \boldsymbol{Z}$.

- $\mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{Y}) \propto \mathcal{L}(Y_1, \dots, Y_n|\boldsymbol{\beta}, \boldsymbol{Z}) \times \mathcal{L}(\boldsymbol{Z}|\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ (\mathbf{1}(y_i = 1)\mathbf{1}(Z_i > 0) + \mathbf{1}(y_i = 0)\mathbf{1}(Z_i < 0)) \, \phi(Z_i; \boldsymbol{x}_i^t \boldsymbol{\beta}, 1) \right\}$ i.e. $Z_1, \dots, Z_n$ are independent given $\boldsymbol{\beta}, \boldsymbol{y}$, and

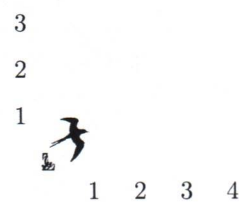
  $$Z_i|\boldsymbol{\beta}, \boldsymbol{y} \sim \begin{cases} \mathbf{1}(Z_i > 0)\phi(Z_i; \boldsymbol{x}_i^t \boldsymbol{\beta}, 1) & \text{if } y_i = 1 \\ \mathbf{1}(Z_i < 0)\phi(Z_i; \boldsymbol{x}_i^t \boldsymbol{\beta}, 1) & \text{if } y_i = 0. \end{cases}$$

  The last two expressions denote the truncated Gaussian densities with support $(0, +\infty)$ and $(-\infty, 0)$, respectively.

BAYESIAN STATISTICS   A. Guglielmi, I. Bianchini   27.02.2017

Properly justify all your answers.

Exercise 1 A research laboratory needs to study the behavior of migratory birds in Italy; for this reason, it registers the direction of departure of a particular species of swallows on the xy-plane. The random variable of interest, called Z, is the angle that the direction of departure of a bird forms with respect to the x-axis, as displayed in the figure below.



The laboratory measures the direction of departure for n swallows, denoted by $(z_1, \ldots, z_n)$, where $z_i \in (-\pi, \pi)$. We assume the following conditional density for datapoints

$$Z_i | \; \mu, \kappa \overset{\text{i.i.d.}}{\sim} f(z; \mu, \kappa) = \frac{e^{\kappa \cos(z-\mu)}}{2\pi I_0(\kappa)}, \quad z \in (-\pi, \pi), \tag{1}$$

where $\mu \in (-\pi, \pi)$ and $\kappa > 0$. Here $I_0(\kappa)$ is a special function called *modified Bessel function* of order 0, defined as

$$I_0(\kappa) = \sum_{m=0}^{+\infty} \frac{1}{m!\Gamma(m+1)} \left(\frac{\kappa}{2}\right)^{2m}.$$

Observe that (1) is known as the density of the von Mises distribution with parameters μ and κ; μ is a measure of location (it is the mean of the distribution) and κ is a measure of concentration.

1. Compute the likelihood $L(\mu, \kappa; z_1, \ldots, z_n)$ for datapoints $(z_1, \ldots, z_n)$.

For (μ, κ), assume the prior density

$$\pi(\mu, \kappa | \mu_0, c, R_0) \propto \frac{1}{(I_0(\kappa))^c} e^{\kappa R_0 \cos(\mu - \mu_0)} \mathbf{1}_{(-\pi,\pi)}(\mu) \; \mathbf{1}_{(0,+\infty)}(\kappa) \; . \tag{2}$$

where $\mu_0 \in (-\pi, \pi)$, c and $R_0 > 0$.

2. Compute the posterior; is the prior conjugate? If the answer is positive, find the hyperparameters of the posterior, denoted μ_n, c_n and R_n.

   (*Hint*: when deriving the expression for μ_n and R_n, remember that

   $$\cos(x-y) = \cos(x)\cos(y) + \sin(x)\sin(y)$$

   and that $\arctan(x)$ is the inverse function of $\tan(x) = \dfrac{\sin(x)}{\cos(x)}$).

3. The prior hyperparameters can be thought of as representing c observations which **average** direction with respect to the x-axis is μ_0. Three old datapoints for Z are available, recorded in October 2006, and representing the direction of departure of 3 birds (with respect to the x-axis): $\boldsymbol{z}^{old} = (0, \pi/3, -\pi/3)$; derive (c, μ_0) according to this information, assuming R_0 equal to 1.

In October 2016 the scientists recorded the following data, with $n = 5$:

z_i	0	$\pi/3$	$\pi/6$	$\pi/6$	0

Table 1: Datapoints

4. Compute the hyperparameters of the posterior of (μ, κ) derived at point 2., using data as in Table 1.

   (*Hint*: it can be useful to know that $\arctan(0.3566528) = 0.3425892$.)

5. The researchers of the laboratory want to understand if, as they think, it is true that this species migrates from Italy towards Hungary, that is located at an angle of $\pi/9$ with respect to Milano (the origin of our coordinate system). With this aim, they assume a new prior, fixing $\kappa = 1$, and $\mu \sim Unif(-\pi, \pi)$. Define a suitable test, specifying the null and the alternative hypothesis, and compute the Bayes Factor between the two hypotheses. What are your conclusion with available data?

   (*Hint*: $I_0(4.625182) = 19.52219$ and $I_0(1) = 1.266066$; moreover, when computing the denominator of the BF, you could use the same trick used to derive expressions for μ_n, R_n at point 2.)

Exercise 2

1. Compute the marginal law of $(X_1, \dots, X_n)$, a sample from a Dirichlet process P with parameters $\alpha > 0$ and P_0 (a probability measure on $\mathbb{R}$) for every fixed integer $n \geq 1$. Provide the proof.

2. Describe the sampling scheme of the generalized Pòlya urn: denote by $(X_1, \dots, X_n)$ the sequence of the first n draws obtained with this scheme. Derive the corresponding predictive laws $\mathcal{L}(X_n | X_{n-1}, \dots, X_1)$ for every fixed $n \geq 2$ and $\mathcal{L}(X_1)$.

3. Is there any relationship between the marginal law computed at point 1. and the marginal law derived at point 2.?

4. Illustrate the stick-breaking construction of the Dirichlet process P (on $\mathbb{R}$).

Solution of EX. 1

1. Since observations are independent and identically distributed, the likelihood is

$$L(\mu, \kappa; z_1, \dots, z_n) = \prod_{i=1}^{n} \frac{e^{\kappa \cos(z_i - \mu)}}{2\pi I_0(\kappa)} = \frac{e^{\kappa \sum_{i=1}^{n} \cos(z_i - \mu)}}{(2\pi I_0(\kappa))^n}$$

$$= \frac{\exp\left(\kappa \sum_{i=1}^{n} \cos(z_i)\cos(\mu) + \kappa \sum_{i=1}^{n} \sin(z_i)\sin(\mu)\right)}{(2\pi I_0(\kappa))^n}, \quad z_i \in (-\pi, \pi), i = 1, \dots, n.$$

2. Thanks to Bayes' theorem, the posterior is proportional to the likelihood times the prior:

$$\pi(\mu,\kappa|z_1,\ldots,z_n) \propto (I_0(\kappa))^{-(c+n)} \exp\left(\kappa R_0\cos(\mu)\cos(\mu_0)+\kappa R_0\sin(\mu)\sin(\mu_0)\right)$$
$$\times \exp\left(\kappa\cos(\mu)\sum_{i=1}^n \cos(z_i)+\kappa\sin(\mu)\sum_{i=1}^n\sin(z_i)\right) I_{(-\pi,\pi)}(\mu)I_{(0,\infty)}(\kappa)$$
$$= (I_0(\kappa))^{-(c+n)}\exp\left\{\kappa\left((R_0\cos\mu_0+\sum\cos z_i)\cos\mu+(R_0\sin\mu_0+\sum\sin z_i)\sin\mu\right)\right\}$$
$$\times I_{(-\pi,\pi)}(\mu)I_{(0,\infty)}(\kappa)$$

This should be of the same family as the prior if this expression is equal to

$$\pi(\mu,\kappa|z_1,\ldots,z_n) \propto (I_0(\kappa))^{-(c_n)}\exp\left(\kappa R_n\cos(\mu-\mu_n)\right) I_{(-\pi,\pi)}(\mu)I_{(0,\infty)}(\kappa)$$
$$= (I_0(\kappa))^{-(c+n)}\exp\left(\kappa R_n\cos(\mu_n)\cos\mu+\kappa R_n\sin(\mu_n)\sin\mu\right) I_{(-\pi,\pi)}(\mu)I_{(0,\infty)}(\kappa)$$

where we used the hint in the text. Therefore, the prior is conjugate to the likelihood if $c_n = c+n$ and R_n, μ_n satisfy the following equation:

$$\begin{cases} R_n\cos(\mu_n) = R_0\cos(\mu_0)+\sum_{i=1}^n\cos(z_i) \\ R_n\sin(\mu_n) = R_0\sin(\mu_0)+\sum_{i=1}^n\sin(z_i)\end{cases}$$

so that

$$\tan(\mu_n) = \frac{\sin(\mu_n)}{\cos(\mu_n)} = \frac{R_0\sin(\mu_0)+\sum_{i=1}^n\sin(z_i)}{R_0\cos(\mu_0)+\sum_{i=1}^n\cos(z_i)}$$

and finally

$$\mu_n = \arctan\left(\frac{R_0\sin(\mu_0)+\sum_{i=1}^n\sin(z_i)}{R_0\cos(\mu_0)+\sum_{i=1}^n\cos(z_i)}\right)$$

and

$$R_n = \frac{R_0\sin(\mu_0)+\sum_{i=1}^n\sin(z_i)}{\sin(\mu_n)}$$

or equivalently

$$R_n = \sqrt{R_0^2+(\sum_{i=1}^n\sin(z_i))^2+(\sum_{i=1}^n\cos(z_i))^2+2R_0(\cos(\mu_0)\sum_{i=1}^n\cos(z_i)+\sin(\mu_0)\sum_{i=1}^n\sin(z_i))}.$$

3. From the interpretation of the parameters of the prior, we have $c=3$ and μ_0 is the average direction of the old sample, $\mathbf{z}^{old}$, namely $\mu_0 = 0$.

4. By simply updating the parameters with the observations, we obtain $c_n = 5+3 = 8$, $\mu_n = \arctan(1.866025/(1+4.232051)) = atan(0.3566528) = 0.3425892 \simeq \pi/9$ and $R_n = 5.554855$.

5. The null and alternative hypothesis to be tested are

$$H_0:\ \mu=\bar{\mu}:=\pi/9 \quad vs \quad H_1:\ \mu\neq\bar{\mu}:=\pi/9$$

where the dimension of H_0 and H_1 is different (punctual hypothesis).

The Bayes Factor is then given by the ratio of the marginal distribution under the two hypotheses, H_0 and H_1.

The marginal under H_1 is given by

$$m_1(z_1,\ldots,z_n) = \int_{-\pi}^{\pi}\prod_{i=1}^n\left(\frac{e^{\cos(z_i-\mu)}}{2\pi I_0(1)}\right)\frac{1}{2\pi}d\mu = \int_{-\pi}^{\pi}\frac{e^{\sum_{i=1}^n\cos(z_i-\mu)}}{(2\pi I_0(1))^n}\frac{1}{2\pi}d\mu$$

$$= \frac{1}{(2\pi)^{n+1}(I_0(1))^n} \int_{-\pi}^{\pi} \exp\left(\sum_{i=1}^{n} \cos(z_i)\cos(\mu) + \sum_{i=1}^{n} \sin(z_i)\sin(\mu) \right) d\mu.$$

We repeat the same trick as in point 2., where

$$\begin{cases} R_1 \cos(\mu_1) = \sum_{i=1}^n \cos(z_i) \\ R_1 \sin(\mu_1) = \sum_{i=1}^n \sin(z_i) \end{cases}$$

and

$$\mu_1 = \arctan\left(\frac{\sum_{i=1}^n \sin(z_i)}{\sum_{i=1}^n \cos(z_i)} \right), \qquad R_1 = \frac{\sum_{i=1}^n \sin(z_i)}{\sin(\mu_n)}.$$

Therefore,

$$\begin{aligned} m_1(z_1, \dots, z_n) &= \frac{1}{(2\pi)^{n+1}(I_0(1))^n} \int_{-\pi}^{\pi} \exp\left(R_1 \cos\mu_1 \cos\mu + R_1 \sin\mu_1 \sin\mu\right) d\mu \\ &= \frac{1}{(2\pi)^{n+1}(I_0(1))^n} \int_{-\pi}^{\pi} e^{R_1 \cos(\mu-\mu_1)} d\mu \\ &= \frac{2\pi I_0(R_1)}{(2\pi)^{n+1}(I_0(1))^n} \int_{-\pi}^{\pi} \frac{e^{R_1 \cos(\mu-\mu_1)}}{2\pi I_0(R_1)} d\mu = \frac{I_0(R_1)}{(2\pi)^n (I_0(1))^n}, \end{aligned}$$

since the last integrand function is a density. Plugging the data in the formula, we obtain $m_1(z_1, \dots, z_n) = 0.0006128407$.

On the other hand, the marginal under H_0 is given by:

$$m_0(z_1, \dots, z_n) = f(z_1, \dots, z_n | \frac{\pi}{9}, 1) = \frac{\exp(\sum_{i=1}^n \cos(z_i - \pi/9))}{(2\pi)^n (I_0(1))^n} = \frac{\exp(4.615045)}{(2\pi)^n (I_0(1))^n}.$$

In conclusion, we have

$$BF_{01} = \frac{m_0(z_1, \dots, z_n)}{m_1(z_1, \dots, z_n)} = \frac{\exp(\sum_{i=1}^n \cos(z_i - \pi/9))}{I_0(R_1)} = 5.173211$$

and $2\log(BF_{01}) \simeq 3.287$, suggesting evidence in favor of H_0.

∎

Solution of EX. 2 We provide a sketch of the solution: for more details, see the textbook.

1. The marginal distribution is given by

$$\mathcal{L}(dX_1, \dots, dX_n) = \prod_{i=1}^{n} \left(\frac{\alpha P_0(dX_i) + \sum_{j=1}^{i-1} \delta_{X_j}(dX_i)}{\alpha + i - 1} \right)$$

with the convention that the element in the brackets is P_0 when $i = 1$. and can be obtained by marginalizing out the random measure P.

2. The generalized Polya urn is a sampling scheme where we start with an urn filled with α black balls. Then we proceed as follows: (i) Each time we need an observation, we draw a ball from the urn; (ii) If the ball is black, we generate a new (non-black) color from the distribution P_0, label a new ball this color, drop the new ball into the urn along with the ball we drew, and return the color we generated. (iii) Otherwise, label a new ball with the color of the ball we drew, drop the new ball into the urn along with the ball we drew, and return the color we observed. The predictive law, for each n, is

$$\mathcal{L}(dX_n | X_1, \dots, X_{n-1}) = \frac{\alpha P_0(dX_n) + \sum_{j=1}^{n-1} \delta_{X_j}(dX_n)}{\alpha + n - 1}.$$

3. The resulting distribution over labels with the sampling scheme derived in 2. is the same as the distribution over values in a Dirichlet process.

4. Let P_0 be a probability measure on $\mathbb{R}$, and $\alpha > 0$. Consider two independent families of r.v.'s, $\{Y_k, k = 1, 2, \ldots\}$ and $\{\theta_k, k = 1, 2, \ldots\}$, where

$$Y_k \overset{\text{i.i.d.}}{\sim} Beta(1, \alpha), \qquad \theta_k \overset{\text{i.i.d.}}{\sim} P_0 \; .$$

Define

$$w_1 = Y_1, \qquad w_k = Y_k \prod_{j=1}^{k-1} (1 - Y_j) \; k = 2, 3, \ldots \; .$$

Then

$$P := \sum_{k=1}^{+\infty} w_k \delta_{\theta_k} \sim DP(\alpha, P_0).$$

Note that

$$\sum_{k=1}^{+\infty} w_k = 1 \text{ a.s.},$$

and that each weight correspond to an infinite number of pieces of a unit-length stick, each piece length proportional to $Beta(1, \alpha)$ random variables.

■

BAYESIAN STATISTICS A. Guglielmi, I. Bianchini 03.02.2017

Properly justify all your answers.

Exercise 1 Consider the following autoregressive model of order 1 for a sequence of random variables $(X_t)_{t\in\mathbb{N}}$, where:

$$\begin{aligned} &X_0 = 0 \\ &X_t = \rho X_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,\sigma^2) \quad t = 1,2,\ldots, \end{aligned} \tag{1}$$

where $\rho \in \mathbb{R}, \sigma^2 > 0$. We observe T values of the sequence, i.e., $x_{1:T} = (x_1, x_2, \ldots, x_T)$.

1. Derive the likelihood, $L(\rho, \sigma^2 | x_{1:T})$, that is the conditional distribution $\mathcal{L}(X_1, \ldots, X_T | \rho, \sigma^2)$.
2. As a prior for (ρ, σ^2), assume

   $$\pi(\rho,\sigma^2) = \pi(\rho|\sigma^2)\pi(\sigma^2) = \mathcal{N}(\rho; \mu_0, \kappa\sigma^2) \times inv - gamma(\sigma^2; \alpha, \beta), \ \mu_0 \in \mathbb{R}, \kappa, \alpha, \beta > 0.$$

   Show that this prior is conjugate to (1), and compute the posterior; in particular, derive the hyperparameters of the posterior distribution.

   Hint: it can be useful to keep in mind that the posterior factorizes in the same way as the prior:

   $$\pi(\rho, \sigma^2 | x_{1:T}) = \pi(\rho|\sigma^2, x_{1:T})\pi(\sigma^2|x_{1:T}).$$

   Thus, it would be easier to identify $\pi(\rho|\sigma^2, x_{1:T})$ first, and then derive $\pi(\sigma^2|x_{1:T})$.
3. Fix α and β such that a priori $\mathbb{E}(\sigma^2) = 0.1$ and $\mathrm{Var}(\sigma^2) = 1$.

We observe the following data, where $T = 5$,

t	1	2	3	4	5
x_t	-0.4	-0.22	-0.1	0.05	0

Table 1: Data points

and $\sum_{t=1}^5 x_t^2 = 0.2209$, $\sum_{t=1}^5 x_t x_{t-1} = 0.105$, $\sum_{t=1}^5 x_{t-1}^2 = 0.2209$.

4. For data in Table 1, assuming $\mu_0 = 0$ and $\kappa = 10$, compute the hyperparameters of the posterior derived at point 2.
5. Compute the predictive density $m_{X_6|X_{1:5}}(x; x_{1:5})$ of the next observation, using the sample in Table 1. Do not forget the normalizing constant.
6. Suppose now that σ^2 is fixed and equal to 1. Assume, as a prior for ρ,

   $$\rho \sim \pi_2(\rho) = gamma(\rho; a = 2, b = 1).$$

   Derive the corresponding posterior, up to a normalizing constant, with data in Table 1. Since the distribution does not belong to a known parametric family, use the acceptance-rejection algorithm to sample from the posterior. Describe all the steps of the procedure, with the assumptions you may need.

Exercise 2 We consider data $\{(y_i, \delta_i),\ i = 1, \dots, n\}$, where $\delta_i = 0$ if the $i - th$ observation is right censored and $\delta_i = 1$ if it is not. The value y_i is a realization from the r.v. $\min(T_i, C_i)$, where C_i is the censoring time and T_i is the time-to-event for subject i. We assume

$$T_i|\theta \overset{\text{ind.}}{\sim} f_i(t|\theta), \quad i = 1, \dots, n$$

where θ is a parameter vector and $f_i(\cdot|\theta)$ represents a density. Derive the likelihood $L(\theta|(y_1, \delta_1), \dots, (y_n, \delta_n))$, and prove your formula, specifying the assumptions.

Hint: assume that T_i and C_i are discrete random variables.

Solution of EX. 1

1. The conditional distribution of $(X_1, \dots, X_T)$ given (ρ, σ^2) is

$$\mathcal{L}(X_{1:T}|\rho, \sigma^2) = \prod_{i=1}^{T} \mathcal{L}(X_t|X_{t-1}, \rho, \sigma^2) = \prod_{i=1}^{T} \mathcal{N}(X_t|\rho X_{t-1}, \sigma^2)$$

Therefore, the likelihood is

$$\begin{aligned}
L(\rho, \sigma^2|x_{1:T}) &= \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x_t - \rho x_{t-1})^2\right\} = (2\pi\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(x_t - \rho x_{t-1})^2\right\} \\
&= (2\pi)^{-T/2}(\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}(\rho^2 \sum_{t=1}^{T} x_{t-1}^2 - 2\rho \sum_{t=1}^{T} x_{t-1}x_t + \sum_{t=1}^{T} x_t^2)\right\}, \ \rho \in \mathbb{R}, \sigma^2 > 0,
\end{aligned}$$

where $x_0 = 0$.

2. Applying Bayes' theorem, we have

$$\begin{aligned}
\pi(\rho, \sigma^2|x_{1:T}) &\propto L(\rho, \sigma^2|x_{1:T})\pi(\rho, \sigma^2) \\
&\propto (\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}(\rho^2 \sum_{t=1}^{T} x_{t-1}^2 - 2\rho \sum_{t=1}^{T} x_{t-1}x_t + \sum_{t=1}^{T} x_t^2)\right\} \\
&\quad \times (\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\kappa\sigma^2}(\rho - \mu_0)^2\right) (\sigma^2)^{-\alpha-1} \exp\left(-\beta/\sigma^2\right) \mathbf{1}_{(0,+\infty)}(\sigma^2) \\
&\propto (\sigma^2)^{-T/2-1/2-\alpha-1} \exp\left\{-\frac{1}{2\sigma^2}(\rho^2 \sum_{t=1}^{T} x_{t-1}^2 - 2\rho \sum_{t=1}^{T} x_{t-1}x_t + \sum_{t=1}^{T} x_t^2)\right\} \\
&\quad \times \exp\left\{-\frac{1}{2\kappa\sigma^2}(\rho^2 + \mu_0^2 - 2\mu_0\rho)\right\} e^{-\frac{\beta}{\sigma^2}} \mathbf{1}_{(0,+\infty)}(\sigma^2)
\end{aligned}$$

If we collect all the terms in the exponential, we got:

$$\begin{aligned}
&-\frac{1}{2\sigma^2}\left(\rho^2 \sum_{t=1}^{T} x_{t-1}^2 - 2\rho \sum_{t=1}^{T} x_{t-1}x_t + \sum_{t=1}^{T} x_t^2 + \frac{1}{\kappa}\rho^2 + \frac{\mu_0^2}{\kappa} - 2\frac{\mu_0}{\kappa}\rho\right) \\
&= -\frac{1}{2\sigma^2}(\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\kappa}) \left(\rho - \frac{\sum_{t=1}^{T} x_{t-1}x_t + \frac{\mu_0}{\kappa}}{\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\kappa}}\right)^2 - \frac{1}{2\sigma^2}\left(\sum_{t=1}^{T} x_t^2 + \frac{\mu_0^2}{\kappa} - \frac{(\sum_{t=1}^{T} x_{t-1}x_t + \frac{\mu_0}{\kappa})^2}{\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\kappa}}\right) \\
&= -\frac{1}{2\sigma^2}(\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\kappa}) \left(\left(\rho - \frac{\sum_{t=1}^{T} x_{t-1}x_t + \frac{\mu_0}{\kappa}}{\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\kappa}}\right)^2 + \frac{\sum_{t=1}^{T} x_t^2 + \frac{\mu_0^2}{\kappa}}{\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\kappa}} - \left(\frac{\sum_{t=1}^{T} x_{t-1}x_t + \frac{\mu_0}{\kappa}}{\sum_{t=1}^{T} x_{t-1}^2 + \frac{1}{\kappa}}\right)^2\right)
\end{aligned}$$

Hence

$$\pi(\rho,\sigma^2|x_{1:T}) \propto (\sigma^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(\sum_{t=1}^{T} x_{t-1}^2+\frac{1}{\kappa})(\rho-\mu_T)^2\right\}$$
$$\times(\sigma^2)^{-T/2-\alpha-1}\mathrm{e}^{-\frac{\beta}{\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{t=1}^{T}x_t^2+\frac{\mu_0^2}{\kappa}-\mu_T^2\right)\right\}\mathbf{1}_{(0,+\infty)}(\sigma^2).$$

The first factor is the kernel of a Gaussian density, i.e.,

$$\pi(\rho|x_{1:T},\sigma^2)=\mathcal{N}(\rho;\mu_T,\kappa_T\sigma^2),$$

while the second corresponds to

$$\pi(\rho|x_{1:T},\sigma^2)=inv-gamma(\sigma^2;\alpha_T,\beta_T),$$

where

- $\mu_T=\kappa_T\left(\sum_{t=1}^T x_t x_{t-1}+\frac{\mu_0}{\kappa}\right)=\dfrac{\sum_{t=1}^T x_t x_{t-1}+\frac{\mu_0}{\kappa}}{\sum_{t=1}^T x_{t-1}^2+\frac{1}{\kappa}}$
- $\kappa_T=\left(\sum_{t=1}^T x_{t-1}^2+\frac{1}{\kappa}\right)^{-1}$
- $\alpha_T=\alpha+T/2$
- $\beta_T=\beta+\frac{1}{2}\left(\sum_{t=1}^T x_t^2+\mu_0^2/\kappa-(\mu_T)^2/\kappa_T\right)$

3. We choose α and β such that

$$\mathbb{E}(\rho)=\frac{\beta}{\alpha-1}=0.1 \qquad \mathrm{Var}(\rho)=\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}=1.$$

We got $\beta=\dfrac{\alpha-1}{10}$ and hence $\alpha-2=\dfrac{1}{100}$, so that $\alpha=2.01$, $\beta=0.101$.

4. The parameters in the posterior are: $\mu_T=0.3272047$, $\kappa_T=3.116236$, $\alpha_T=4.51$ and $\beta_T=0.1942718$.

5. First of all, $\mathcal{L}(X_{T+1}|\rho,\sigma^2,x_T)$ does not depend on ρ, since $x_T=0$; hence

$$m_{X_{T+1}|x_{1:T}}(x)=\mathcal{L}(X_{T+1}|x_{1:T})=\int_{\mathbb{R}}\int_0^{+\infty}\mathcal{L}(X_{T+1}|\rho,\sigma^2,x_T)\pi(d\rho,d\sigma^2|x_{1:T})$$
$$=\int_{\mathbb{R}}\int_0^{+\infty}\mathcal{N}(x;0,\sigma^2)\mathcal{N}(d\rho;\mu_T,\kappa_T\sigma^2)\times inv-gamma(d\sigma^2;\alpha_T,\beta_T)$$
$$=\int_0^{+\infty}\mathcal{N}(x;0,\sigma^2)\times inv-gamma(d\sigma^2;\alpha_T,\beta_T)=\int_0^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\frac{\beta_T^{\alpha_T}}{\Gamma(\alpha_T)}(\sigma^2)^{-\alpha_T-1}\mathrm{e}^{-\beta_T/\sigma^2}d\sigma^2$$
$$=\frac{\beta_T{}^{\alpha_T}}{\Gamma(\alpha_T)\sqrt{2\pi}}\int_0^{+\infty}(\sigma^2)^{-\alpha_T-1/2-1}\mathrm{e}^{-\frac{1}{\sigma^2}(\beta_T+\frac{x^2}{2})}d\sigma^2=\frac{\beta_T{}^{\alpha_T}}{\Gamma(\alpha_T)\sqrt{2\pi}}\frac{\Gamma(\alpha_T+1/2)}{\left(\beta_T+\frac{x^2}{2}\right)^{\alpha_T+1/2}}.$$

6. The posterior of ρ is now proportional to

$$\pi(\rho|x_{1:T}) \propto \exp\left(-\frac{1}{2}(\rho^2\sum_{t=1}^{T}x_{t-1}^2 - 2\rho\sum_{t=1}^{T}x_{t-1}x_t)\right)\frac{b^a}{\Gamma(a)}\rho^{a-1}\mathrm{e}^{-b\rho}\mathbf{1}_{(0,+\infty)}(\sigma^2).$$

We may apply the acceptance-rejection method if we are able to find a constant M and a density $g(x)$ such that

$$f(x) \le Mg(x), \forall x.$$

Note that f can be defined also up to a normalizing constant. In our specific case we have

$$f(x) = \exp\left(-\frac{1}{2}(\rho^2\sum_{t=1}^{T}x_{t-1}^2 - 2\rho\sum_{t=1}^{T}x_{t-1}x_t)\right)\rho^{a-1}\mathrm{e}^{-b\rho} \le \exp\left(-\rho(b-\sum_{t=1}^{T}x_{t-1}x_t)\right)\rho^{a-1}, \quad \rho > 0$$

so that we define $g(\cdot)$ as the density of the $gamma(a, b - \sum_{t=1}^{T} x_{t-1}x_t)$ that is well defined if

$$b - \sum_{t=1}^{T} x_{t-1}x_t > 0.$$

The condition is satisfied in this case, since $1 - 0.105 > 0$.

In summary, we have that $g(x)$ is the density of a gamma distribution with parameters $\left(a, b - \sum_{t=1}^{T} x_{t-1}x_t\right)$. The constant M is therefore $\Gamma(a)/(b - \sum_{t=1}^{T} x_{t-1}x_t)^a$.

The acceptance-rejection algorithm works as follows:

(a) sample $Y \sim gamma(a = 2, b - \sum_{t=1}^{T} x_{t-1}x_t = 0.895)$ and compute the acceptance rate $r = \frac{f(Y)}{Mg(Y)}$;

(b) generate $U \sim \mathcal{U}((0,1))$; if $U \le r$, accept Y, otherwise go to (a).

Note that it is always possible to generate from $Z \sim gamma(2,1)$ distribution (generating two r.v.s iid from the exponential distribution with parameter equal to 1, and this is done via the inverse distribution method, and then summing them), and then $Y = \dfrac{1}{0.895}Z \sim gamma(2, 0.895)$.

∎

Solution of EX. 2 We have

$$\mathcal{L}((y_1,\delta_1),\ldots,(y_n,\delta_n)|\theta) \propto \prod_{i=1}^{n}\left((f_i(y_i|\theta))^{\delta_i}(S_i(y_i|\theta))^{1-\delta_i}\right)$$

meaning that if $\delta_i = 1$ the contribution to the likelihood is given by the density, otherwise we only know that $T_i > C_i = y_i$, so the contribution is $\mathbb{P}(T_i > C_i = y_i|\theta) = S_i(y_i|\theta)$. This formula may be proved when:

(i) T_i and C_i are independent (cond. to all parameters their marginal distributions depend on),
(ii) the distribution of C_i does not depend on the parameter θ,
(iii) all the subjects in the sample are independent, conditionally to all parameters.

For more details, see *Christensen et al (2011), Chapter 12.* ∎

BAYESIAN STATISTICS A. Guglielmi and I. Bianchini 13.09.2016

Properly justify all your answers.

Exercise 1 Assume the following model for data $X_1, X_2, \dots, X_n$:

$$X_1, \dots, X_n | p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2 \overset{iid}{\sim} p\mathcal{N}(\mu_1, \sigma_1^2) + (1-p)\mathcal{N}(\mu_2, \sigma_2^2) \tag{1}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance $\sigma^2 > 0$, and $p \in (0,1)$. The prior for $(p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ is such that $\pi(p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = \pi(p)\pi(\mu_1, \sigma_1^2)\pi(\mu_2, \sigma_2^2)$ and

$$\begin{aligned} &\mu_j | \sigma_j^2 \sim \mathcal{N}(\xi_j, \sigma_j^2 / n_j), \quad \sigma_j^2 \sim inv-gamma\left(\frac{\nu_j}{2}, \frac{s_j^2}{2}\right), \quad j = 1,2 \\ &p \sim Beta(a,b) \end{aligned} \tag{2}$$

where $(a, b, \xi_1, \xi_2, n_1, n_2)$ are fixed hyperparameters, $a, b > 0$, $\xi_j \in \mathbb{R}$, $n_j, \nu_j, s_j^2 > 0$ for $j = 1,2$. The conditional distribution in (1) can be equivalently written through a missing variable representation, by introducing n latent variables $Z_1, \dots, Z_n$ such that, conditioning on $Z_1, \dots, Z_n$ and the parameters, the r.v.s $X_1, \dots, X_n$ are independent and, for all $i = 1, \dots, n$,

$$\begin{aligned} &X_i | Z_i, p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2 \overset{ind}{\sim} \mathcal{N}(\mu_{Z_i}, \sigma_{Z_i}^2) \\ &Z_i | p \overset{iid}{\sim} modBe(p), \end{aligned}$$

where $modBe(p)$ denotes a binary random variable assuming value 1 with probability p and value 2 with probability $1-p$.

Define $\mathbf{x} := (x_1, \dots, x_n)$ the vector of data, the set $\mathcal{C}_j := \{i \in \{1, 2, \dots, n\} : Z_i = j\}$ and $l_j := \#\mathcal{C}_j$, $j = 1,2$, with $l_1 + l_2 = n$. In order to simulate from the posterior of model (1)-(2) through a Gibbs sampler algorithm, we augment the parameter space with $\mathbf{Z} = (Z_1, \dots, Z_n)$. The posterior of interest is then

$$\pi(\mathbf{Z}, p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2 | \mathbf{x}). \tag{3}$$

1. Derive the joint distribution $\mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ of data, latent variables and parameter $\boldsymbol{\theta} := (p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$.

2. Show that

$$\begin{aligned} &p | \mathbf{Z}, \mathbf{x} \sim Beta(a + l_1, b + l_2) \\ &\mu_j | \sigma_j^2, \mathbf{Z}, \mathbf{x} \sim \mathcal{N}\left(\xi_j(\mathbf{Z}), \frac{\sigma_j^2}{n_j + l_j}\right), \ \sigma_j^2 | \mathbf{Z}, \mathbf{x} \sim inv-gamma\left(\frac{\nu_j + l_j}{2}, \frac{s_j^2(\mathbf{Z})}{2}\right), \quad j = 1,2 \end{aligned} \tag{4}$$

   where $\xi_j(\mathbf{Z}) = \dfrac{n_j \xi_j + \sum_{i \in \mathcal{C}_j} x_i}{n_j + l_j}$ and $s_j^2(\mathbf{Z}) = \left(s_j^2 + \sum_{i \in \mathcal{C}_j} x_i^2 + n_j \xi_j^2 - (n_j + l_j)(\xi_j(\mathbf{Z}))^2\right)$.

   (**Hint:** the inverse-gamma density with parameter (α, β) is $(\beta^\alpha / \Gamma(\alpha)) u^{-\alpha-1} e^{-\beta/u} \mathbb{I}_{(0,+\infty)}(u)$, $\alpha, \beta > 0$)

3. Build a Gibbs sampler to simulate from the posterior (3).

4. From the MCMC sample $\left(\mathbf{z}^{(k)}, p^{(k)}, \mu_1^{(k)}, \sigma_1^{2(k)}, \mu_2^{(k)}, \sigma_2^{2(k)}\right)$, $k = 1, \dots, G$, obtained at point 3., derive a Monte Carlo estimate of the $n \times n$ posterior incidence matrix I, with entries

$$I_{ll} := 1, \quad I_{lm} := \mathbb{P}(Z_l = Z_m | \mathbf{x}), \ l \neq m \ \ l, m = 1, \dots, n.$$

5. With the aim of comparing two models, $\mathbf{M_0}$ and $\mathbf{M_1}$, which are both a simplified version of model (1)-(2):

$$\boldsymbol{M}_0 : X_1|p, \mu_1, \mu_2 \sim p\,\mathcal{N}(x; \mu_1, 1) + (1-p)\mathcal{N}(x; \mu_2, 1)$$
$$\mu_j \stackrel{\text{ind.}}{\sim} \mathcal{N}(\xi_j, 1/n_j) \quad j = 1,2, \quad p \sim Beta(a,b), \quad p \perp (\mu_1, \mu_2)$$

$$\boldsymbol{M}_1 : X_1|\mu_1 \sim \mathcal{N}(x; \mu_1, 1), \qquad \mu_1 \sim \mathcal{N}(\xi_1, 1/n_1),$$

compute the Bayes factor BF_{01} of model $\boldsymbol{M}_0$ vs $\boldsymbol{M}_1$ when $n = 1$ and $x_1 = 0.5$. In this case, assume $(\xi_1, \xi_2, n_1, n_2) = (-1, 0.5, 2, 2)$ and set a, b in $\mathbf{M_0}$, such that

$$\mathrm{E}_\pi(p) = 60\% \qquad \text{and} \qquad \mathrm{Var}_\pi(p) = 0.2.$$

Which model do you prefer? Why?

(**Hint:** remember that, if $p \sim Beta(a,b)$, then $\mathrm{E}[p^k] = \dfrac{a+k-1}{a+b+k-1}\mathrm{E}[p^{k-1}]$ for all positive integers k)

Solution of EX. 1

1. The joint distribution of $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ is

$$\begin{aligned}\mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) &= \mathcal{L}(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta})\mathcal{L}(\boldsymbol{Z}|\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta})\\ &= \prod_{i=1}^n \mathcal{N}(\mu_{z_i}, \sigma^2_{z_i}) \prod_{i=1}^n (p\mathbb{I}_{Z_i=1} + (1-p)\mathbb{I}_{Z_i=2})\, \pi(p)\pi(\mu_1|\sigma_1^2)\pi(\sigma_1^2)\pi(\mu_2|\sigma_2^2)\pi(\sigma_2^2).\end{aligned}$$

2. All the conditional distributions to be checked are proportional to the joint distribution described at point 1. Hence:

$$\begin{aligned}\mathcal{L}(p|\mathbf{Z}, \mathbf{x}) \propto \mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) &\propto \prod_{i=1}^n (p\mathbb{I}_{Z_i=1} + (1-p)\mathbb{I}_{Z_i=2})\, \pi(p)\\ &\propto p^{l_1}(1-p)^{l_2} p^{a-1}(1-p)^{b-1}\mathbb{I}_{(0,1)}(p) \propto p^{a+l_1-1}(1-p)^{b+l_2-1}\mathbb{I}_{(0,1)}(p)\end{aligned}$$

that is the $Beta(a+l_1, b+l_2)$ distribution. Analogously, if $j = 1,2$, then

$$\begin{aligned}\mathcal{L}(\mu_j|\sigma_j^2, \mathbf{Z}, \mathbf{x}) \propto \mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) &\propto \prod_{i=1}^n \mathcal{N}(x_i; \mu_{Z_i}, \sigma^2_{Z_i})\pi(\mu_j|\sigma_j^2)\\ &\propto \prod_{i\in C_j} \frac{1}{\sqrt{2\pi\sigma^2_{Z_i}}} \mathrm{e}^{-\frac{(x_i-\mu_{Z_i})^2}{2\sigma^2_{Z_i}}} \frac{1}{\sqrt{2\pi\sigma_j^2/n_j}} \mathrm{e}^{-(\mu_j-\xi_j)^2/(2\sigma_j^2)/n_j} = \mathcal{N}\left(\xi_j(\mathbf{Z}), \frac{\sigma_j^2}{n_j+l_j}\right)\end{aligned}$$

and

$$\begin{aligned}\mathcal{L}(\sigma_j^2|\mathbf{Z}, \mathbf{x}) \propto \mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) &\propto \prod_{i\in C_j} \mathcal{N}(x_i; \mu_{Z_i}, \sigma^2_{Z_i})\ \pi(\sigma_j^2)\\ &\propto \prod_{i\in C_j} \frac{1}{\sqrt{2\pi\sigma^2_{Z_i}}} \mathrm{e}^{-\frac{(x_i-\mu_{Z_i})^2}{2\sigma^2_{Z_i}}} \left(\frac{1}{\sigma_j^2}\right)^{(\nu_j/2)+1} \mathrm{e}^{-\frac{s_j^2}{2\sigma_j^2}}\mathbb{I}_{(0,+\infty)}(\sigma_j^2)\\ &= inv-gamma\left(\frac{\nu_j+l_j}{2}, \frac{s_j^2(\mathbf{Z})}{2}\right),\end{aligned}$$

since this is the usual update of the Gaussian-inverse gamma conjugate model, where the parameters are updated with respect to the observations in each group. See, for instance, the solution of the exam exercise, on 24 September 2014.
Notation $\mathcal{N}(x, \mu, \sigma^2)$ denotes the Gaussian density with mean μ and variance σ^2, evaluated at x.

3. First, we derive all the full-conditionals of the augmented parameter space where $(\boldsymbol{Z}, \boldsymbol{\theta}) = (\boldsymbol{Z}, p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ lives.
   (a) $\mathcal{L}(\mathbf{Z}|\mathbf{x}, \boldsymbol{\theta}) \propto \mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) = \prod_{i=1}^n \left(\mathcal{N}(x_i; \mu_{z_i}, \sigma_{z_i}^2)\,(p\mathbb{I}_{Z_i=1} + (1-p)\mathbb{I}_{Z_i=2})\right)$, i.e. it is the product of the n full conditionals $\mathcal{L}(Z_i|\mathbf{x}, \boldsymbol{\theta})$, which is the discrete distribution defined here below:
$$\mathbb{P}(Z_i = 1|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p) \propto p\mathcal{N}(x_i; \mu_1, \sigma_1^2)$$
$$\mathbb{P}(Z_i = 2|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p) \propto (1-p)\mathcal{N}(x_i; \mu_2, \sigma_2^2)$$
   (b) $\mathcal{L}(p|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \mathbf{Z}, \mathbf{x}) \propto \mathcal{L}(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) = \mathcal{L}(p|\mathbf{Z}, \mathbf{x})$; this is the $Beta(a+l_1, b+l_2)$ distribution derived at point 1.
   (c) $\mathcal{L}(\mu_1|\sigma_1^2, \mu_2, \sigma_2^2, p, \mathbf{Z}, \mathbf{x}) = \mathcal{L}(\mu_1|\sigma_1^2, \mathbf{Z}, \mathbf{x})$ and $\mathcal{L}(\sigma_1^2|\mu_1, \mu_2, \sigma_2^2, p, \mathbf{Z}, \mathbf{x}) = \mathcal{L}(\sigma_1^2|\mathbf{Z}, \mathbf{x})$ were calculated at point 2. An analogous result holds for the full conditionals of μ_2 and σ_2^2.

   The Gibbs sampler is built by repeatedly sampling from these distributions at the n-th step, conditioning on values of the parameters obtained at the $n-1$-th step.

4. We have that
$$I_{lm} = \mathbb{P}(Z_l = Z_m|\mathbf{x}) = \mathbb{E}\left(\mathbb{I}(Z_l = Z_m)|\mathbf{x}\right) \simeq \frac{1}{G}\sum_{g=1}^{G} \mathbb{I}(z_l^{(g)} = z_m^{(g)}) \quad \text{for any } l \neq m.$$

5. First, we derive (a, b) according to prior information:
$$\mathrm{E}_\pi(p) = \frac{a}{a+b} = 0.6 \qquad \text{and} \qquad \mathrm{Var}_\pi(p) = \frac{ab}{(a+b)^2(a+b+1)} = 0.2$$
$$a = (\mathrm{E}_\pi(p)(1-\mathrm{E}_\pi(p)) - \mathrm{Var}_\pi(p)) \times \left(\mathrm{Var}_\pi(p)\left(1 + \frac{1-\mathrm{E}_\pi(p)}{\mathrm{E}_\pi(p)}\right)\right)^{-1}$$
and
$$b = \frac{1-\mathrm{E}_\pi(p)}{\mathrm{E}_\pi(p)}a.$$
Therefore we have $a = 0.12$ and $b = 0.08$.
Now, we need to compute the marginal distributions under the two models:
$$m_0(x_1) = \int\int\int (p\mathcal{N}(x_1; \mu_1, 1) + (1-p)\mathcal{N}(x_1; \mu_2, 1))\, beta(dp; a, b)\mathcal{N}(d\mu_1; \xi_1, 1/n_1)\mathcal{N}(d\mu_2; \xi_2, 1/n_2)$$
$$= I_1(x_1) + I_2(x_1)$$
where
$$I_1(x_1) = \int_{\mathbb{R}}\int_0^1 \frac{1}{\sqrt{2\pi}} p \exp\left(-\frac{1}{2}(x_1-\mu_1)^2\right)\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}dp d\mu_1$$
$$= \sqrt{\frac{n_1}{1+n_1}}\frac{a}{a+b}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x_1^2 + n_1\xi_1^2 - \frac{(x_1+n_1\xi_1)^2}{1+n_1})\right) = \frac{a}{a+b}\mathcal{N}(x_1; \xi_1, \frac{n_1+1}{n_1})$$

and

$$I_2(x_1) = \sqrt{\frac{n_2}{1+n_2}} \frac{b}{a+b} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_1^2 + n_2\xi_2^2 - \frac{(x_1+n_2\xi_2)^2}{1+n_2})\right) = \frac{b}{a+b}\mathcal{N}(x_1;\xi_2,\frac{n_2+1}{n_2}).$$

Similarly, we have

$$m_1(x_1) = \int \mathcal{N}(x_1;\mu_1,1)\mathcal{N}(d\mu_1;\xi_1,1/n_1) = \sqrt{\frac{n_1}{1+n_1}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_1^2 + n_1\xi_1^2 - \frac{(x_1+n_1\xi_1)^2}{1+n_1})\right)$$
$$= \mathcal{N}(x_1;\xi_1,\frac{n_1+1}{n_1})$$

Plugging the given values in the formulas, we obtain

$$m_0(x_1) = 0.2226138 \quad m_1(x_1) = 0.1538663 \quad BF_{01} = m_0(x_1)/m_1(x_1) = 1.4468.$$

Alternatively

$$BF_{01} = 0.6 + 0.4\frac{\frac{1}{\sqrt{2\pi\frac{3}{2}}}e^{-(x_1-(1/2))^2/3}}{\frac{1}{\sqrt{2\pi\frac{3}{2}}}e^{-(x_1+1)^2/3}} = 0.6 + 0.4e^{x_1+\frac{1}{4}} = 0.6 + 0.4e^{0.75} \simeq 1.4468.$$

Finally, we have that $2\log BF_{01} \simeq 0.7387$, showing weak evidence in favour of $\boldsymbol{M}_0$.

∎

BAYESIAN STATISTICS A. Guglielmi and I. Bianchini 06.07.2016

Properly justify all your answers.

Exercise 1 Let $(Y_1, \dots, Y_k)$ be a random vector of dimension $k \geq 2$ with multinomial density with parameters $(n, \boldsymbol{\theta})$, where $\boldsymbol{\theta} := (\theta_1, \dots, \theta_k)$, i.e.

$$\mathbb{P}(Y_1 = y_1, \dots, Y_k = y_k) = \frac{n!}{y_1! \dots y_k!} \theta_1^{y_1} \dots \theta_k^{y_k}, \quad y_1, \dots, y_k \in \{0, 1, \dots, n\}, \ y_1 + y_2 + \dots + y_k = n, \quad (1)$$

with $0 < \theta_j < 1$, $j = 1, 2, \dots, k$ and $\theta_1 + \theta_2 + \dots + \theta_k = 1$.

1. Find a conjugate prior for the parameter $\boldsymbol{\theta}$. In order to fix notation, use the parameterization $(a_1, \dots, a_k)$ such that $\mathbb{E}(\theta_j) = a_j/\alpha$ and $\alpha = \sum_{j=1}^k a_j$. Write explicitly the density of the parameters $(\theta_1, \dots, \theta_{k-1})$, and the range of the hyperparameters.
2. Compute the posterior of $\boldsymbol{\theta}$, having observed $(y_1, \dots, y_k)$ from (1).
3. Derive the marginal distribution of the vector $(\theta_1, 1 - \theta_1)$. Which distribution do you recognize?
4. Which is the marginal distribution for θ_1 obtained from point 3.? Compute the maximum a-posteriori (MAP) estimate of θ_1.

The table below shows the voting intention of a sample of 740 Austrian voters at 2016 presidential election, after an important public debate. Assume the data modeled through a multinomial distribution as in (1) with $k = 4$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_4)$. As a prior for $\boldsymbol{\theta}$, consider the conjugate prior found at point 1. with hyperparameters (a_1, a_2, a_3, a_4).

Candidates	Van der Bellen	Hofer	Hundstorfer	No preference	Total
No. of voters	288	332	101	19	740

Table 1: Election poll after the debate.

5. Declared votes in a poll **before** the debate, over 250 interviewees, were: 99 for Van der Bellen, 87 for Hofer, 50 for Hundstorfer and the remaining 14 did not express a preference. Use this information to fix the hyperparameters (a_1, a_2, a_3, a_4), when the total mass $\sum_{j=1} a_j$ equal to 100: in particular, assume that the marginal prior mean of each parameter θ_j coincides with the corresponding empirical estimates from the poll before the debate.
6. Compute the MAP estimate of the probability of voting Van der Bellen, using hyperparameters obtained at point 5. and data in Table 1.
7. Since you are not extremely confident on the choice of hyperparameter a_1, you assume this value as a random variable, i.e.

$$a_1 \sim gamma(\alpha, \beta), \quad \text{with } \alpha > 0 \text{ and } \beta > 0,$$

and fix $(a_2, a_3, a_4) = (0.6, 0.3, 0.1)$ (differently from what you assumed at point 5. and 6.). Propose a Gibbs sampler to sample from the posterior distribution $\mathcal{L}(a_1, \theta_1, \theta_2, \theta_3 | y_1, \dots, y_4)$.

(**Hint:** remember that $\int_D \theta_1^{a_1-1} \theta_2^{a_2-1} \theta_3^{a_3-1} (1 - \theta_1 - \theta_2 - \theta_3)^{a_4-1} d\theta_1 d\theta_2 d\theta_3 = \frac{\prod_{j=1}^4 \Gamma(a_j)}{\Gamma(\sum_{j=1}^4 a_j)}$ where D is the support of the Dirichlet(a_1, a_2, a_3, a_4) distribution.)

Exercise 2 Let P be a Dirichlet process on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, i.e. $P \sim DP(a, P_0)$, where $a > 0$ and P_0 is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Derive $\mathrm{E}(P(A))$ and $\mathrm{Var}(P(A))$ for every A in $\mathcal{B}(\mathbb{R})$. Moreover, compute $\mathrm{Cov}(P(A), P(B))$ for every A, B in $\mathcal{B}(\mathbb{R})$ with $A \cap B = \varnothing$.

Solution of EX. 1

1. A conjugate prior $\boldsymbol{\theta} := (\theta_1, \dots, \theta_k)$ in case of multinomial likelihood is the Dirichlet$(a_1, a_2, \dots, a_k)$ distribution, with $a_j > 0$ for $j = 1, \dots, k$; note that this prior for $\boldsymbol{\theta}$ is degenerate on $\mathbb{R}^k$, but, under $\theta_k = 1 - \theta_1 - \dots - \theta_{k-1}$, $\theta_1, \dots, \theta_{k-1}$ has prior density (w.r.t. the Lebesgue measure on $\mathbb{R}^{k-1}$):

$$\pi(\theta_1, \dots, \theta_{k-1}|a_1, \dots, a_k) = \frac{1}{B(\mathbf{a})} \prod_{i=1}^{k-1} \theta_i^{a_i - 1}(1 - \sum_{j=1}^{k-1} \theta_j)^{a_k - 1}, 0 < \theta_1 + \dots + \theta_{k-1} < 1,\ \theta_j \in (0,1) \text{ for all } j$$

and $B(\mathbf{a}) = \dfrac{\prod_{j=1}^k \Gamma(a_j)}{\Gamma(\sum_{j=1}^k a_j)}$. In this way, $\mathbb{E}(\theta_i) = a_i / (\sum_1^k a_j)$ for $i = 1, \dots, k$, as requested.

2. The posterior distribution is proportional to the likelihood times the prior:

$$\pi(\theta_1, \dots, \theta_{k-1}|y_1, \dots, y_k) \propto \theta_1^{y_1 + a_1 - 1} \theta_2^{y_2 + a_2 - 1} \dots \theta_k^{y_k + a_k - 1},\ 0 < \theta_1 + \dots + \theta_{k-1} < 1,\ \theta_j \in (0,1) \text{ for all } j.$$

This is again a Dirichlet$(a_1 + y_1, a_2 + y_2, \dots, a_k + y_k)$ distribution, where the parameters have been updated with the observations.

3. By the aggregation property, we have $(\theta_1, 1 - \theta_1) \sim$ Dirichlet$(a_1, \sum_{j=2}^k a_j)$. Be careful: this distribution is NOT absolute continuous on $\mathbb{R}^2$, i.e. it does not have a density in $\mathbb{R}^2$.

4. On the other hand, the prior distribution of θ_1 has a density on $\mathbb{R}$, since it is Beta with parameters $(a_1, \sum_{j=2}^k a_j)$, by definition of the 2-dimensional Dirichlet distribution.

   MAP: if we compute the argmax of the beta density with general positive parameters (α, β), we have to solve the following equation:

$$\frac{df(\theta)}{d\theta} \propto \frac{d}{d\theta}\left[\theta^{\alpha-1}(1-\theta)^{\beta-1}\right] = (\alpha-1)\theta^{\alpha-2}(1-\theta)^{\beta-1} - (\beta-1)\theta^{\alpha-1}(1-\theta)^{\beta-2} \geq 0,$$

   i.e.

$$\alpha - 1 - \theta(\alpha - 1 + \beta - 1) \geq 0 \iff \theta \leq \frac{\alpha - 1}{\alpha + \beta - 2} =: \hat{\theta} \text{ when } \alpha, \beta > 1.$$

   In this case, this value is a minimum. Summing up, the mode of $\theta \sim$Beta(α, β) is equal to $\hat{\theta}$ for $\alpha, \beta > 1$.

   Therefore, under the same argument as in 3. and here above, a posteriori $\theta_1 \sim$Beta$(a_1 + y_1, \sum_2^k (a_j + y_j))$=Beta$(a_1 + y_1, M - a_1 + n - y_1)$, where $M = \sum_1^k a_j$. When all $a_i + y_i > 1$; we obtain that the MAP of θ_1 is

$$\hat{\theta}_1 = \frac{a_1 + y_1 - 1}{a_1 + y_1 + M - a_1 + n - y_1 - 2} = \frac{a_1 + y_1 - 1}{M + n - 2}.$$

5. In this case $M = \sum_1^4 a_j = 100$; assuming that $\mathrm{E}(\theta_j) = a_j / M =: f_j$, $j = 1, 2, 3, 4$, where f_j denotes the empirical estimate of the mean, we have:

$$a_1 = \frac{99}{250} \times M, \quad a_2 = \frac{87}{250} \times M, \quad a_3 = \frac{50}{250} \times M, \quad a_4 = \frac{14}{250} \times M,$$

   so that $(a_1, a_2, a_3, a_4) = (39.6, 34.8, 20, 5.6)$.

6. The MAP estimate is equal to

$$\hat{\theta}_1 = \frac{39.6 + 288 - 1}{100 + 740 - 2} = \frac{326.6}{838} \simeq 0.3997.$$

7. In order to derive a Gibbs sampler, we need the full-conditionals, that are proportional to the joint law of data and parameters:

$$\mathcal{L}(a_1, \theta_1, \dots, \theta_3, y_1, \dots, y_4)$$
$$\propto \frac{n!}{y_1! \dots y_4!} \theta_1^{y_1} \dots \theta_3^{y_3} (1-\theta_1-\theta_2-\theta_3)^{y_4} \times \frac{1}{B(\mathbf{a})} \theta_1^{a_1-1} \dots \theta_3^{a_3-1} (1-\theta_1-\theta_2-\theta_3)^{a_4-1} \times a_1^{\alpha-1} e^{-\beta a_1}$$
$$\propto \theta_1^{y_1} \dots \theta_3^{y_3} (1-\theta_1-\theta_2-\theta_3)^{y_4} \times \frac{\Gamma(a_1+1)}{\Gamma(a_1)} \theta_1^{a_1-1} \dots \theta_3^{a_3-1} (1-\theta_1-\theta_2-\theta_3)^{a_4-1} \times a_1^{\alpha-1} e^{-\beta a_1}$$

where $0 < \theta_1 + \theta_2 + \theta_3 < 1$, $0 < \theta_1, \theta_2, \theta_3 < 1$, $a_1 > 0$, and $y_1, \dots, y_k$ in $\{0, 1, \dots, n\}$, $y_1 + y_2 + \dots + y_k = n$. Then, the full-conditionals are:

- $(\theta_1, \theta_2, \theta_3) | \ a_1, y_1, \dots, y_4 \ \sim \mathit{Dirichlet}(a_1 + y_1, 0.6 + y_2, 0.3 + y_3, 0.1 + y_4)$;
- $a_1 | \theta_1, \theta_2, \theta_3, y_1, \dots, y_4 \ \propto \ \frac{\Gamma(a_1+1)}{\Gamma(a_1)} \theta_1^{a_1-1} \times a_1^{\alpha-1} e^{-\beta a_1} = e^{log(\theta_1) a_1} a_1^{\alpha+1-1} e^{-\beta a_1} \mathbf{1}_{(0,+\infty)}(a_1)$, i.e.

$$a_1 | (\theta_1, \theta_2, \theta_3, y_1, \dots, y_4) \sim \mathit{gamma}(\alpha + 1, \beta - \mathit{log}(\theta_1)).$$

∎

Solution of EX. 2 For the properties of the Dirichlet process, we know that, for every $A \in \mathcal{B}(\mathbb{R})$, $(P(A), P(A^c)) \sim \mathit{Dirichlet}(aP_0(A), aP_0(A^c))$, so that

$$P(A) \sim \mathit{Beta}(aP_0(A), a(1 - P_0(A)))$$

since $P(A^c) = 1 - P(A)$. Therefore,

$$\mathbb{E}(P(A)) = \frac{aP_0(A)}{aP_0(A) + a(1 - P_0(A))} = P_0(A)$$

and

$$\begin{aligned} \mathrm{Var}(P(A)) &= \frac{aP_0(A) \times a(1 - P_0(A))}{(aP_0(A) + a(1 - P_0(A)))^2 \times (aP_0(A) + a(1 - P_0(A)) + 1)} \\ &= \frac{a^2 P_0(A)(1 - P_0(A))}{a^2(a+1)} = \frac{P_0(A)(1 - P_0(A))}{a+1}. \end{aligned}$$

Consider now $A, B \in \mathcal{B}(\mathbb{R})$, $A \cap B = \varnothing$: the partition generated by A and B is $A, B, A^c \cap B^c$, and, by the definition of the Dirichlet process,

$$(P(A), P(B), P(A^c \cap B^c)) \sim \mathit{Dirichlet}(aP_0(A), aP_0(B), aP_0(A^c \cap B^c)).$$

Then, $\mathrm{Cov}(P(A), P(B)) = \mathbb{E}(P(A)P(B)) - P_0(A)P_0(B)$. We have:

$$\begin{aligned} \mathbb{E}(P(A)P(B)) &= \int_D xy \frac{\Gamma(a) x^{aP_0(A)-1} y^{aP_0(B)-1}}{\Gamma(aP_0(A))\Gamma(aP_0(B))\Gamma(aP_0(A^c \cap B^c))} (1 - x - y)^{aP_0(A^c \cap B^c)-1} dxdy \\ &= \frac{\Gamma(a)}{\Gamma(aP_0(A))\Gamma(aP_0(B))\Gamma(aP_0(A^c \cap B^c))} \frac{\Gamma(aP_0(A)+1)\Gamma(aP_0(B)+1)\Gamma(aP_0(A^c \cap B^c))}{\Gamma(a+2)} \\ &= \frac{aP_0(A) \ aP_0(B)}{a(a+1)} = \frac{a}{a+1} P_0(A) \ P_0(B). \end{aligned}$$

where $D = \{x, y > 0 : 0 < x + y < 1\}$. Consequently:

$$\mathrm{Cov}(P(A), P(B)) = P_0(A)P_0(B) \left(\frac{a}{a+1} - 1 \right) = -\frac{P_0(A)P_0(B)}{a+1}.$$

∎

BAYESIAN STATISTICS A. Guglielmi, I. Bianchini 01.03.2016

Properly justify all your answers.

Exercise 1 Consider the following model for binary data: $Y_1, Y_2, \ldots, Y_{n_1}, Y_{n_1+1}, \ldots, Y_{n_1+n_2}$ independent conditionally on θ, and

$$Y_1, Y_2, \ldots, Y_{n_1}|\theta \overset{i.i.d.}{\sim} Be(\theta), \qquad Y_{n_1+1}, Y_{n_1+2}, \ldots, Y_{n_1+n_2}|\theta \overset{i.i.d.}{\sim} Be(\theta^2),$$

where Be stands for the Bernoulli distribution, with $\theta \in (0,1)$, and n_1, n_2 are two positive integers. Define $X = \sum_{i=1}^{n_1} Y_i$, $Z = \sum_{i=1}^{n_2} Y_{n_1+i}$ and $n = n_1 + n_2$. We assume a beta density with parameters a and b as the prior for θ, i.e. $\theta \sim beta(a,b)$, with $a, b > 0$.

1. Given observations $X = x$ and $Z = z$, where $x \in \{0, 1, \ldots, n_1\}$, $z \in \{0, 1, \ldots, n_2\}$, show that the posterior of θ is a mixture of beta distributions. In particular, compute the posterior density as a mixture of beta densities, and calculate weights and parameters of the component densities. You will find that the weights are functions of beta functions as

$$B(s,t) := \int_0^1 u^{s-1}(1-u)^{t-1} du = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}, \quad s, t > 0.$$

Hint: use the formula $(\theta + 1)^m = \sum_{j=0}^{m} \binom{m}{j} \theta^j$, for any non-negative integer m. Notation $a_j^* = a + x + 2z + j, b^* = b + n - (x + z)$ might be useful.

Two different laboratories use the model described above to make inference; they perform the same experiment, measuring some binary variables. In particular, they assume the same value for n_1 and get the same value for $X = x$, where $x \in \{0, \ldots, n_1\}$. As far as Z is concerned, we have that $n_2 = 1$ for both laboratories, while the observed value is different: the first laboratory gets $z_1 = 0$, but the second obtains $z_2 = 1$.

2. Compute the posterior distributions $\pi_1(\cdot; x, z_1, n_1, n_2)$ and $\pi_2(\cdot; x, z_2, n_1, n_2)$ of the first and second laboratory, respectively. In particular, evaluate the normalized weights of both posterior density mixtures, getting rid of the beta functions in the expression derived at point 1.

3. The two laboratories share the same prior information. They know that marginal mean and variance of Y_1 and Y_{n_1+1}, first and last observation, satisfy the following constraints: the mean of Y_1 is 4 times the mean of Y_{n_1+1}, while the variance of Y_1 is 64/19 times the variance of Y_{n_1+1}. Compute the hyperparameters of the prior, a and b.

Hint: the $k-th$ moment of the $beta(a,b)$ random variable is

$$\frac{a(a+1)\ldots(a+k-1)}{(a+b)(a+b+1)\ldots(a+b+k-1)}.$$

4. Assume now that $x = 2$ and $n_1 = 10$. Compute posterior means and variances of θ obtained by the two laboratories.

Exercise 2 For data $y_1, \ldots, y_n$, assume they are conditionally independent, given the model parameter θ, so that the likelihood is

$$f(y_1, \ldots, y_n|\theta) = \prod_{i=1}^{n} f_i(y_i|\theta),$$

where $f_i(\cdot|\theta)$ denotes the conditional density of Y_i. Assume that a priori $\theta \sim \pi(\theta)$, where π denotes a density on Θ.

Give the definition of Conditional Predictive Ordinate for the $i-th$ datum y_i (CPO_i) and the Log-Pseudo Marginal Likelihood (LPML) of the sample $(y_1, \ldots, y_n)$ for the model considered here. In addition, explain how these indexes can be used to assess the goodness of fit. Use notation $\mathbf{y}_{-i}$ for the whole vector of data but the i-th point.

Suppose now to have a sample (obtained via a MCMC algorithm) $(\theta^{(1)}, \ldots, \theta^{(M)})$ from the posterior distribution $\pi(\theta|y_1, \ldots, y_n)$: show how CPO_i, $i = 1, \ldots, n$, and thus LPML, can be estimated using this sample.

Solution of EX. 1

1. The likelihood is

$$f(x, z|\theta) \propto \theta^x(1-\theta)^{n_1-x}(\theta^2)^z(1-\theta^2)^{n_2-z} = \theta^{x+2z}(1-\theta)^{n-x-z}(1+\theta)^{n_2-z}$$
$$= \sum_{j=0}^{n_2-z} \binom{n_2-z}{j} \theta^{x+2z+j}(1-\theta)^{n-x-z}$$

where $\theta \in (0,1)$, $x = 0, 1, \ldots, n_1$ and $z = 0, 1, \ldots, n_2$. We used formula $(\theta+1)^{n_2-z} = \sum_{j=0}^{n_2-z} \binom{n_2-z}{j}\theta^j$. If the prior is $\theta \sim beta(a, b)$, then

$$\pi(\theta|x, y, n_1, n_2) \propto \sum_{j=0}^{n_2-z} \binom{n_2-z}{j} \theta^{x+2z+j+a-1}(1-\theta)^{n-x-z+b-1}\mathbf{1}_{(0,1)}(\theta).$$

Denoting $a_j^* := a + x + 2z + j$, with $j = 0, \ldots, n_2 - z$, and $b^* := b + n - (x + z)$, we can write

$$\pi(\theta|x, z, n_1, n_2) \propto \sum_{j=0}^{n_2-z} \binom{n_2-z}{j} B(a_j^*, b^*)\frac{1}{B(a_j^*, b^*)}\theta^{x+2z+j+a-1}(1-\theta)^{n-x-z+b-1}\mathbf{1}_{(0,1)}(\theta)$$
$$= \sum_{j=0}^{n_2-z} \binom{n_2-z}{j} B(a_j^*, b^*)k_B(\theta; a_j^*, b^*),$$

where $k_B(\cdot, a, b)$ stands for the density of a beta distribution with parameters a, b. It is then obvious that

$$\pi(\theta|x, z, n_1, n_2) = \frac{1}{K}\sum_{j=0}^{n_2-z} \binom{n_2-z}{j} B(a_j^*, b^*)k_B(\theta; a_j^*, b^*)$$

where $K = \sum_{j=0}^{n_2-z} \binom{n_2-z}{j} B(a_j^*, b^*)$, so that the weight of the j-th component is

$$w_j = \frac{1}{K}\binom{n_2-z}{j} B(a_j^*, b^*).$$

2. Here, we exploit the results obtained above. For the first laboratory, we know that $z_1 = 0$, so $n_2 - z_1 = 1 - 0 = 1$, therefore the posterior is a mixture of two beta densities. More in detail, given $a_0^* = a + x$, $a_1^* = a + x + 1$, $b^* = b + n - x = b + n_1 + 1 - x$, we have:

$$\pi_1(\theta|x, 0, n_1, 1) \propto B(a_0^*, b^*)k_B(\theta; a_0^*, b^*) + B(a_1^*, b^*)k_B(\theta; a_1^*, b^*).$$

Now observe that

$$B(a_1^*, b^*) = \frac{\Gamma(x+1+a)\Gamma(n-x+b)}{\Gamma(a+b+n+1)} = \frac{(x+a)\Gamma(x+a)\Gamma(n-x+b)}{(a+b+n)\Gamma(a+b+n)} = \frac{x+a}{a+b+n}B(a_0^*, b^*)$$

and consequently

$$\pi_1(\theta|x, z, n_1, n_2) \propto k_B(\theta; a+x, b+n-x) + \frac{x+a}{a+b+n}k_B(\theta; a+x+1, b+n-x).$$

The weights are proportional to 1 and $\frac{x+a}{a+b+n}$, that sum up to $\frac{2a+b+n+x}{a+b+n}$. Normalizing these weights we obtain

$$\begin{aligned}\pi_1(\theta|x, z, n_1, n_2) &= \frac{a+b+n}{x+2a+b+n}k_B(\theta; a+x, b+n-x) \\ &\quad + \frac{x+a}{x+2a+b+n}k_B(\theta; a+x+1, b+n-x).\end{aligned}$$

The same result can be obtained directly from point 1., noting that

$$K = B(a_0^*, b^*) + B(a_1^*, b^*) = B(a_0^*, b^*)\left(1 + \frac{a+x}{a+b+n}\right),$$

and that the weight of the first component of the mixture $k_B(\theta; a_0^*, b^*)$ is

$$\frac{B(a_0^*, b^*)}{B(a_0^*, b^*) + B(a_1^*, b^*)} = \frac{B(a_0^*, b^*)}{B(a_0^*, b^*)\left(1 + \frac{a+x}{a+b+n}\right)} = \frac{a+b+n}{x+2a+b+n}.$$

On the other hand, for the second laboratory, we have that $z_1 = 1$ and $n_2 - z_1 = 1 - 1 = 0$: therefore, the posterior is a mixture with just one component. In particular, $a_0^* = a + x + 2$ and $b^* = b + n - x - 1 = b + n_1 - x$, so that:

$$\pi_2(\theta, x, z_2, n_1, n_2) = k_B(\theta; a+x+2, b+n_1-x).$$

3. Let us fix the hyperparameters of the prior using the available information:

$$\mathrm{E}(Y_1) = 4\,\mathrm{E}(Y_{n_1+1}) \quad \mathrm{Var}(Y_1) = \frac{64}{19}\mathrm{Var}(Y_{n_1+1}).$$

All Y_i's are marginally Bernoulli distributed, where the parameter is the mean of the correspondent random variable, namely

$$\mathrm{E}(Y_1) = \mathrm{E}(\mathrm{E}(Y_1|\theta)) = \frac{a}{a+b}, \qquad \mathrm{E}(Y_{n_1+1}) = \mathrm{E}(\mathrm{E}(Y_{n_1+1}|\theta) = \mathrm{E}(\theta^2) = \frac{a(a+1)}{(a+b)(a+b+1)}.$$

Therefore, $Y_1 \sim Be(a/a+b)$ and $Y_{n_1+1} \sim Be\left(\frac{a(a+1)}{(a+b)(a+b+1)}\right)$. Remember that a Bernoulli variable of parameter p has variance equal to $p(1-p)$; then the information we have can be represented as

$$\begin{aligned}\frac{a}{a+b} &= 4\frac{a(a+1)}{(a+b)(a+b+1)} \\ \frac{a}{a+b}\left(1 - \frac{a}{a+b}\right) &= \frac{64}{19}\frac{a(a+1)}{(a+b)(a+b+1)}\left(1 - \frac{a(a+1)}{(a+b)(a+b+1)}\right),\end{aligned}$$

so that

$$\begin{aligned} \frac{a+1}{a+b+1} &= \frac{1}{4} \\ \left(1-\frac{a}{a+b}\right) &= \frac{64}{19}\times\frac{1}{4}\left(1-\frac{1}{4}\frac{a}{a+b}\right) \end{aligned}$$

From the second equation we have

$$\frac{a}{a+b}=\frac{1}{5}\Rightarrow b=4a,$$

and substituting into the first one gives

$$\frac{a+1}{5a+1}=\frac{1}{4}, \quad \Longrightarrow \quad a=3, b=12.$$

4. For laboratory 1, we know that $a=3$, $b=12$, $n_1=10$, $x=2$, $z_1=0$, $n_2=1$ and $n=11$. From point 2., we have that $\pi_1(\cdot;x,z_1,n_1,n_2)$ is a linear combination of densities

$$k_B(\theta;3+2,12+11-2) \text{ and } k_B(\theta;3+2+1,12+11-2),$$

with weights $\frac{a+b+n}{2a+b+n+x}=\frac{26}{31}$ and $\frac{a+x}{2a+b+n+x}=\frac{5}{31}$, respectively. Summing up,

$$\pi_1(\cdot;x,z_1,n_1,n_2)=\frac{26}{31}k_B(\theta;5,21)+\frac{5}{31}k_B(\theta;6,21).$$

Moreover,

$$\begin{aligned} \mathrm{E}_{\pi_1}\left(\theta|x,z_1,n_1,n_2\right) &= \frac{26}{31}\times\frac{5}{26}+\frac{5}{31}\times\frac{6}{27}\simeq 0.1971326 \\ \mathrm{E}_{\pi_1}\left(\theta^2|x,z_1,n_1,n_2\right) &= \frac{26}{31}\times\frac{5\times 6}{26\times 27}+\frac{5}{31}\times\frac{6\times 7}{27\times 28}\simeq 0.04480287 \\ \mathrm{Var}_{\pi_1}\left(\theta|x,z_1,n_1,n_2\right) &= 0.04480287-(0.1971326)^2\simeq 0.005941608. \end{aligned}$$

For laboratory 2 the posterior is a single beta distribution with parameters $a+x+2=7$, $b+n_1-x=20$. So we have:

$$\begin{aligned} \mathrm{E}_{\Pi_2}(\theta|x,z_2,n_1,n_2) &= \frac{7}{27}=0.2592593 \\ \mathrm{Var}(\theta|x,z_2,n_1,n_2) &= \frac{7\times 20}{27^2\times 28}=\frac{35}{5103}\simeq 0.006858711. \end{aligned}$$

∎

Solution of EX. 2 See Section 4.9.2 of the book of Christensen et al, *Bayesian ideas and data analysis*, 2011. ∎

BAYESIAN STATISTICS A. Guglielmi and I. Bianchini 17.02.2016

Properly justify all your answers.

Exercise 1 Let $X_1, X_2, \ldots, X_n$ be (conditionally) independent and identically distributed random variables following the Pareto density with parameters $c > 0$ and $\alpha > 0$:

$$f(x;\alpha) = \frac{\alpha c^\alpha}{x^{\alpha+1}} \mathbb{I}_{(c,+\infty)}(x), \quad \alpha > 0, \tag{1}$$

where c is fixed and α is random.

1. Find the density of the random variable $Y = \log\left(\dfrac{X}{c}\right)$, where X has density (1). Which distribution do you recognize? Denote by $g(y;\alpha)$ the density of Y.

Consider now $Y_1, Y_2, \ldots, Y_n$, where $Y_i = log\left(\dfrac{X_i}{c}\right)$, $i = 1, \ldots, n$; each Y_i represents the deviation of observation X_i from the reference value c in the logarithmic scale; obviously, we have that

$$Y_1, Y_2, \ldots, Y_n | \alpha \overset{iid}{\sim} g(\cdot;\alpha). \tag{2}$$

2. Find a prior density $\pi(\alpha)$ for α that is conjugate to likelihood (2). Compute the hyperparameters of the posterior of α, after having observed a sample $(y_1, \ldots, y_n)$. In order to fix notation, let $\pi(\alpha) = \pi(\alpha; a, b)$, where a, b are positive hyperparameters such that $\mathbb{E}(\alpha) = a/b$ and $Var(\alpha) = a/b^2$.

The Pareto is a heavy-tailed distribution, and it is widely used in modeling extreme values. Here, we are interested in analysing the behavior of the **maximum sea level in Venice**, where the annual maximums in the last 12 years are:

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Max Level x_i (cm)	123	122	120	114	96	125	124	120	132	166	134	138

Table 1: Data $(x_1, \ldots, x_{12})$.

Here $c = 80$ cm is the threshold over which water floods the city. For data in Table 1, transformed as $y_i = log\left(\frac{x_i}{c}\right)$ for all i, assume model (2) with prior density $\pi(\alpha; a, b)$ as derived at point 2 .

3. An old book reports few historic values:

Year	1935	1936	1937	1938	1939	1940
Max Level x_i^{old} (cm)	103	81	121	116	115	147

Table 2: Old measurements $(x_1^{old}, \ldots, x_6^{old})$.

Use data in Table 2, suitably transformed, to fix prior hyperparameters a, b, of $\pi(\alpha; a, b)$, adopting the marginal method as follows. Fix $a = 5$ and for $\mathbf{z} = (z_1, \ldots, z_6)$, $z_i = log(x_i^{old}/c)$ for all i, choose $b > 0$ such that

$$b = argmax_b(m(\mathbf{z}; b)),$$

where $m(\mathbf{z})$ is the marginal density under model (2) with prior $\pi(\alpha; a, b)$.

4. Compute the posterior distribution of α, i.e. evaluate the hyperparameters of the posterior of α, given $y_1, \ldots, y_{12}$.

In climatology, the **return value** corresponding to T years is the value x_T such that

$$\mathbb{P}(X > x_T|\alpha) = \frac{1}{T},$$

i.e. x_T is the threshold which is exceeded on average once in T years.

5. Calculate the expression of the return value x_T when $X|\alpha \sim Pareto(c, \alpha)$ as in (1).

6. Provide the posterior expected value of $h(\alpha) =: \log\left(\frac{x_T}{c}\right)$, corresponding to $T = 100$ years, given $y_1, \ldots, y_{12}$ obtained from Table 1.

7. Evaluate the conditional predictive probability that during the current year, the maximum sea level exceeds 110cm, given that the greatest value recorded till today is 92 cm, i.e. evaluate

$$\mathbb{P}(X^{new} > 110 \text{ cm}|X^{new} \geq 92 \text{ cm}, y_1, \ldots, y_n).$$

Hint: It is useful to compute $\mathbb{P}(X^{new} > k|y_1, \ldots, y_n)$ first

Exercise 2 Given a sample $(\theta_1, \ldots, \theta_n)$ from a Dirichlet process with parameters $\alpha > 0$ and P_0 (a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$), namely

$$\begin{aligned} \theta_1, \ldots, \theta_n|P &\overset{iid}{\sim} P \\ P|\alpha, P_0 &\sim DP(\alpha, P_0), \end{aligned} \tag{3}$$

write the predictive distribution $\mathcal{L}(\theta_{n+1}|\theta_1, \ldots, \theta_n)$ of the "next" observation θ_{n+1} from model (3) (**generalized Polya-Urn scheme**). Illustrate a strategy to sample from this law, when P_0 is the gamma distribution with parameters (m, b), where $b > 0$ and m is a positive integer.

Solution of EX. 1

1. If $Y = t(X) = \log(X/c)$, then $Y \geq 0$ a.s. since $X \geq c$ a.s.. Therefore,

$$x = t^{-1}(y) = ce^y, \quad \frac{dt^{-1}(y)}{dy} = ce^y$$

yield

$$g(y; \alpha) = f(t^{-1}(y); \alpha)\left|\frac{dt^{-1}(y)}{dy}\right| = \alpha e^{-y\alpha}\mathbb{I}_{(0,+\infty)}(y),$$

that is the density of the exponential distribution with parameter $\alpha > 0$.

Alternatively, for $y > 0$:

$$F_Y(y) = \mathbb{P}(\log(X/c) \leq y) = \mathbb{P}(X \leq ce^y) = \int_c^{ce^y} \frac{\alpha c^\alpha}{x^{\alpha+1}}dx = \alpha c^\alpha \left[\frac{x^{-\alpha}}{\alpha}\right]_{ce^y}^{c} = 1 - e^{-\alpha y};$$

on the other hand, $F_Y(y) = 0$ when $y \leq 0$. This is the distribution function of $Y \sim Exp(\alpha)$.

2. We have that $Y_1, \ldots, Y_n|\alpha \overset{iid}{\sim} Exp(\alpha)$, so that, for $y_1, \ldots, y_n > 0$, the likelihood is

$$\mathcal{L}(\alpha; y_1, \ldots, y_n) = \prod_{i=1}^{n} g(y_i; \alpha) = \alpha^n \mathrm{e}^{-\alpha \sum_{i=1}^{n} y_i} \mathbb{I}_{(0,+\infty)}(\alpha).$$

The conjugate prior is the $gamma(a, b)$ distribution with density

$$\pi(\alpha; a, b) = \frac{b^a}{\Gamma(a)} \alpha^{a-1} \mathrm{e}^{-b\alpha} \mathbb{I}_{(0,+\infty)}(\alpha).$$

The posterior density is

$$\pi(\alpha|y_1, \ldots, y_n) \propto \alpha^{a-1} e^{-b\alpha} \times \alpha^n \mathrm{e}^{-\alpha \sum_{i=1}^{n} y_i} \mathbb{I}_{(0,+\infty)}(\alpha) = \alpha^{(a+n-1)} \mathrm{e}^{-\alpha(b+\sum_{i=1}^{n} y_i)} \mathbb{I}_{(0,+\infty)}(\alpha)$$

so that $\alpha|y_1, \ldots, y_n \sim gamma(a + n, b + \sum_{i=1}^{n} y_i)$.

3. First of all, the marginal density is

$$m(\mathbf{z}; b) = \int_0^{+\infty} \prod_{i=1}^{m} \left(\alpha \mathrm{e}^{-\alpha z_i}\right) \frac{b^a}{\Gamma(a)} \alpha^{a-1} \mathrm{e}^{-b\alpha} d\alpha = \frac{\Gamma(a+m)}{\Gamma(a)} \frac{b^a}{(b + \sum_{i=1}^{m} z_i)^{a+m}},$$

where $m = 6$ and $z_i = \log(x_i^{old}/c) = (0.2527, 0.0124, 0.4138, 0.3716, 0.3629, 0.6084)$.
In order to find $argmax\,(m(\mathbf{z}; b))$, we maximize $\log m(\mathbf{z}; b)$ with respect to b, i.e.

$$\frac{d \log m(\mathbf{z}; b)}{db} = \frac{d}{db}\left(\log\left(\frac{\Gamma(a+m)}{\Gamma(a)}\right) + a \log b - (a+m) \log\left(b + \sum_{i=1}^{m} z_i\right)\right) = \frac{a}{b} - \frac{a+m}{b + \sum_i^m z_i} \geq 0.$$

This implies $b \leq \dfrac{a \sum_{i=1}^{m} z_i}{m}$, so that the $argmax$ is $\hat{b} = \dfrac{a \sum_{i=1}^{m} z_i}{m}$ and the prior is then

$$\alpha \sim gamma(a = 5, b = 1.6848).$$

4. The posterior distribution of α is

$$\alpha|y_1, \ldots, y_n \sim gamma(a^* = a + 12 = 17, b^* = b + \sum_{i}^{n} y_i = 1.6848 + 5.3761 = 7.0609).$$

5. From (1), we have, for $x_T > c$:

$$\mathbb{P}(X > x_T) = \int_{x_T}^{+\infty} \frac{\alpha c^\alpha}{t^{\alpha+1}} dt = \left(\frac{c}{x_T}\right)^\alpha, \tag{4}$$

so that

$$\left(\frac{c}{x_T}\right)^\alpha = \frac{1}{T} \Leftrightarrow x_T = cT^{1/\alpha}.$$

6. We have that $h(\alpha) = \log(x_T/c) = \log(T)/\alpha$; thus

$$\mathbb{E}\left(h(\alpha)|y_1, \ldots, y_n\right) = \mathbb{E}\left(\log(x_T/c)|y_1, \ldots, y_n\right) = \log(T)\mathbb{E}\left(\frac{1}{\alpha}|y_1, \ldots, y_n\right) = \log(T)\frac{b^*}{a^*-1} \cong 2.0323,$$

since $1/\alpha \sim inv - gamma(a^*, b^*)$.

7. First, we note that

$$\mathbb{P}(X^{new} > 110cm | X^{new} \geq 92cm, y_1, \ldots, y_n) = \frac{\mathbb{P}(X^{new} > 110, X^{new} \geq 92 | y_1, \ldots, y_n)}{\mathbb{P}(X^{new} \geq 92 | y_1, \ldots, y_n)}$$
$$= \frac{\mathbb{P}(X^{new} > 110 | y_1, \ldots, y_n)}{\mathbb{P}(X^{new} \geq 92 | y_1, \ldots, y_n)} = \frac{\int_0^\infty \mathbb{P}(X^{new} > 110 | \alpha) \pi(\alpha | \mathbf{y}) d\alpha}{\int_0^\infty \mathbb{P}(X^{new} > 92 | \alpha) \pi(\alpha | \mathbf{y}) d\alpha}.$$

Now, let k be a value greater than c; by (4), we have that

$$\int_0^{+\infty} \mathbb{P}(X^{new} > k | \alpha) \pi(\alpha | \mathbf{y}) d\alpha = \int_0^{+\infty} \left(\frac{c}{k}\right)^\alpha \pi(\alpha | \mathbf{y}) d\alpha$$
$$= \frac{(b^*)^{a^*}}{\Gamma(a^*)} \int_0^{+\infty} \mathrm{e}^{-\alpha \log(k/c)} \alpha^{a^*-1} \mathrm{e}^{-\alpha b^*} d\alpha = \left(\frac{b^*}{b^* + \log(k/c)}\right)^{a^*}.$$

Therefore

$$\mathbb{P}(X^{new} > 110cm | X^{new} \geq 92cm, y_1, \ldots, y_n) = \left(\frac{b^*}{b^* + \log(110/c)}\right)^{a^*} \left(\frac{b^* + \log(92/c)}{b^*}\right)^{a^*}$$
$$= \left(\frac{b^* + \log(92/80)}{b^* + \log(110/80)}\right)^{a^*} = 0.6592036 \cong 0.6592.$$

∎

Solution of EX. 2

It is known that the predictive distribution implied by model (3) is the same as that arising from a generalized Polya Urn sampling scheme, where

$$\mathbb{P}(\theta_{n+1} \in \cdot | \theta_1, \ldots, \theta_n) = \frac{\alpha}{\alpha + n} P_0(\cdot) + \frac{1}{\alpha + n} \sum_{i=1}^n \delta_{\theta_i}(\cdot); \tag{5}$$

(5) is a mixture between an absolutely continuous distribution P_0, i.e. $gamma(m, b)$, and a discrete one, $\sum_{i=1}^n \delta_{\theta_i}(\cdot)$. A strategy to sample from (5) is the following:

1. draw u from $Uniform(0, 1)$;
2. if $0 < u < \frac{\alpha}{\alpha + n}$, then sample θ_{n+1} from the $gamma(m, b)$;
3. if $\frac{\alpha + i - 1}{\alpha + n} < u < \frac{\alpha + i}{\alpha + n}$, then set $\theta_{n+1} = \theta_i$, $i = 1, \ldots, n$.

In order to perform step 2., we remember that, if

$$Y_1, \ldots, Y_m \overset{iid}{\sim} Exp(b)$$

then $X = \sum_{i=1}^m Y_i \sim gamma(m, b)$. So we just need to sample m values $y_1, \ldots, y_m$ from the exponential distribution with parameter b. This can be done by the inverse transformation method, where the distribution function to invert is $F_{Y_1}(y) = (1 - \mathrm{e}^{-by}) \mathbb{I}_{(0, +\infty)}(y)$.

For $k = 1, \ldots, m$:

1. sample u_k from $Uniform(0, 1)$;
2. set $y_k = -\frac{1}{b} \log(1 - u_k)$.

The put $\theta_{n+1} = \sum_1^m y_k$. ∎

STATISTICA BAYESIANA A. Guglielmi, R. Argiento 11.09.2015

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 Scrivere la definizione di *log-pseudo marginal likelihood* (LPML) per un campione osservato $\boldsymbol{y} = (y_1, \ldots, y_n)$. Per uniformare la notazione, si indichi con $\boldsymbol{y}_{-i}$ il vettore di tutte le osservazioni tranne y_i, e con $f_i(y_i; \theta)$, $\theta \in \Theta$, la densità della legge condizionale dell'i-esima osservazione, dato il parametro θ. Scrivere poi una formula per il calcolo dell'LPML a partire da una realizzazione di un algoritmo MCMC $\{\theta_{(k)}, k = 1, 2, \ldots M\}$ dalla posterior di θ, sulla base del vettore completo $\boldsymbol{y}$ delle osservazioni.

Esercizio 2 Si vuole fare inferenza sui tempi di guasto di un sistema formato da 3 componenti connessi in parallelo. Sappiamo che i componenti funzionano in modo indipendente uno dall'altro, che sono tutti dello stesso tipo e che i loro tempi di guasto Y_1, Y_2, Y_3 (espressi in ore) hanno densità beta di parametri $(a, 1)$, con $a > 0$. Si considerano $n = 120$ sistemi, per i quali è noto che $\sum_1^n \log(x_i) = -49.92$, dove x_i rappresenta la durata del sistema i.

1. Qual è la densità della variabile aleatoria X che descrive la durata di un intero sistema?

Siano $X_1, \ldots, X_n$ le variabili aleatorie che rappresentano le durate degli n sistemi. Si assume che esse siano indipendenti, condizionatamente al parametro a, e con distribuzione ricavata al punto 1.

2. Trovare la distribuzione coniugata al modello e scrivere esplicitamente l'aggiornamento degli iperparametri nella prior sulla base dei dati a disposizione.

3. In un esperimento precedente, condotto su dieci sistemi analoghi a quelli in esame, si era registrato che la somma dei logaritmi dei tempi di guasto in ore era stata pari a -1.73. Utilizzare il principio del campione equivalente per fissare gli iperparametri della prior e valutare di conseguenza il loro aggiornamento.

4. Calcolare la probabilità predittiva che un sistema funzioni al più per 30 minuti.

5. Utilizzando una opportuna approssimazione, ricavare l'intervallo HPD (cioè l'intervallo a *highest posterior density*) per a di livello 0.95.

6. Confrontare le ipotesi:

$$H_0\colon\ a \geq 1 \text{ contro } H_1\colon\ a < 1,$$

attraverso il Bayes factor. A quali conclusioni si giunge sulla base dei dati a disposizione? (*Sarà utile sapere che la funzione di ripartizione di una distribuzione* $\chi^2(20)$, *valutata in 10.38, vale 0.0393. Se necessario utilizzare una opportuna approssimazione.*)

Soluzione

1. Poichè il sistema non funziona se e solo se nessuna delle 3 componenti funziona, la durata X del sistema è minore od uguale ad x se e solo se tutte le v.a. Y_i sono minori o uguali ad x. Quindi:

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\max\{Y_1, Y_2, Y_3\} \leq x) = \mathbb{P}(\{Y_1 \leq x\} \cap \cdots \cap \{Y_3 \leq x\}) \\ &= \prod_{i=1}^{3} \mathbb{P}(Y_i \leq x) = (F_Y(x))^3, \end{aligned} \tag{1}$$

visto che Y_1, Y_2, Y_3 sono i.i.d. con f.r. F_Y. In questo caso, poichè $Y_i \sim Beta(a, 1)$, con $a > 0$, allora $f_Y(y; a) = ay^{a-1}(1-y)^{1-1}\mathbf{1}_{(0,1)}(y)$, e, per $y \in (0, 1)$:

$$F_Y(y; a) = \int_0^y au^{a-1}du = y^a.$$

Utilizzando (1) si ottiene:

$$F_X(x; a) = (x^a)^3, \text{ per } x \in (0, 1),$$

e quindi $f_X(x; a) = 3ax^{3a-1}\mathbf{1}_{(0,1)}(x)$, che è la densità della distribuzione beta di parametri $(3a, 1)$, con $a > 0$.

2. La verosimiglianza (legge condizionale dei dati) per $a > 0$ si ricava nel seguente modo:

$$L(\theta; x_1, \ldots, x_{120}) = \prod_{i=1}^{n} f_X(x_i; \theta) = \prod_{i=1}^{120}(3ax_i^{3a-1}) = \left(\prod_1^{120} x_i^{-1}\right) 3^n a^n e^{3a\sum_1^n \log(x_i)}.$$

Non è difficile rendersi conto che tale verosimiglianza è coniugata alla prior $gamma(\alpha, \beta)$, $\alpha, \beta > 0$, per la quale la densità è: $\pi(a) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} a^{\alpha-1}e^{-\beta a}\mathbf{1}_{(0,+\infty)}(a)$. Applicando il teorema di Bayes si ricava facilmente che

$$\pi(a|\boldsymbol{x}) \propto a^{\alpha+n-1}e^{-(\beta-3\sum_1^n \log x_i)a}\ \mathbf{1}_{(0,+\infty)}(a),$$

dove $\boldsymbol{x} = (x_1, \ldots, x_n)$. Quindi la distribuzione a posteriori di a è una $gamma(\alpha_1, \beta_1)$ con $\alpha_1 = \alpha + n$ e $\beta_1 = \beta - 3\sum_1^n \log x_i$.

3. La media a posteriori di θ è

$$\mathrm{E}(a|\boldsymbol{x}) = \left(\frac{\beta - 3\sum_{i=1}^n \log x_i)}{\alpha + n}\right)^{-1} = \left(\frac{\alpha}{\alpha+n}\frac{\beta}{\alpha} + \frac{n}{\alpha+n}\frac{-3\sum_{i=1}^n \log x_i}{n}\right)^{-1}.$$

Quest'ultima relazione, applicando il principio del campione equivalente, ci dice che il parametro α rappresenta il numero di dati di un ipotetico campione $(z_1, \ldots, z_{10})$, mentre β rappresenta la quantità $-3\sum_{i=1}^{10} \log z_i$. Si fissano dunque gli iperparametri della prior con $\alpha = 10$ e $\beta = 5.19$. Pertanto, i valori a posteriori della distribuzione gamma, che esprime la posterior di a, sono $\alpha_1 = 130$, e $\beta_1 = 154.95$.

4. Si tratta di calcolare la probabilità predittiva:

$$\begin{aligned}
\mathbb{P}(X_{n+1} \le 0.5|\boldsymbol{x}) &= \int_0^{+\infty} \mathbb{P}(X_{n+1} \le 0.5|\theta)\pi(\theta|\boldsymbol{x})d\theta = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\int_0^{+\infty}\left(\frac{1}{2}\right)^{3a} a^{\alpha_1-1}e^{-\beta_1 a}da \\
&= \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\int_0^{+\infty} a^{\alpha_1-1}e^{-(\beta_1+3\log 2)a}da = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_1)}{(\beta_1+3\log 2)^{\alpha_1}} \\
&= \left(\frac{\beta_1}{\beta_1+3\log 2}\right)^{\alpha_1} = 0.176752.
\end{aligned}$$

5. Per quanto riguarda la posterior, utilizziamo il TCL per approssimare la distribuzione gamma (con parametro α grande) con la distribuzione gaussiana con pari media e varianza. Ricordiamo che

$$\mathrm{E}(\theta|\boldsymbol{x}) = \frac{\alpha_1}{\beta_1} = 0.8389803 \simeq 0.8390 \quad \mathrm{Var}(\theta|\boldsymbol{x}) = \frac{\alpha_1}{\beta_1^2} = (0.07358344)^2 \simeq (0.0736)^2.$$

Quindi la posterior $\theta \sim gamma(\alpha_1 = 130, \beta_1 = 154.95) \simeq \mathcal{N}(0.8390, (0.0736)^2)$; è immediato ricavare che un intervallpo HPD di livello 0.95 è individuato dagli estremi:

$$0.8390 \pm z_{\alpha/2} \times 0.0736,$$

con $z_{\alpha/2} = z_{0.025} \simeq 1.96$ quantile di coda destra di una gaussiana standard. Pertanto l'IC cercato è

$$(0.7028, 0.9772).$$

Si noti che il valore $a = 1$ non appartiene a tale intervallo.

6. Dobbiamo calcolare il Bayes factor tra le ipotesi H_0 e H_1 come rapporto fra posterior e prior odds, cioè

$$BF_{01} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\mathbb{P}(a \geq 1|\boldsymbol{x}) / \mathbb{P}(a < 1|\boldsymbol{x})}{\mathbb{P}(a \geq 1) / \mathbb{P}(a < 1)}.$$

A priori $\theta \sim gamma(\alpha = 10, \beta = 5.19) = gamma(20/2, 5.19)$.

È facile verificare che $c\theta \sim gamma(20/2, 1/2) = \chi^2(20)$ se $c = 2\beta = 10.38$. Quindi:

$$\mathbb{P}(a < 1) = \mathbb{P}(ca \leq 10.38) = F_{\chi^2(20)}(10.38) = 0.0393.$$

Pertanto il prior odds è

$$\mathbb{P}(a \geq 1) / \mathbb{P}(a < 1) = \frac{1 - F_{\chi^2(20)(10.38)}}{F_{\chi^2(20)(10.38)}} = 0.9607/0.0393 \simeq 24.4453.$$

Per quanto riguarda la posterior, abbiamo visto al punto precedente che si può approssimare con $\mathcal{N}(0.8390, (0.0736)^2)$; allora il posterior odds è pari a:

$$\mathbb{P}(a \leq 1|\boldsymbol{x}) / \mathbb{P}(a > 1|\boldsymbol{x}) \simeq \frac{1 - \Phi\left(\frac{1-0.8390}{0.0736}\right)}{\Phi\left(\frac{1-0.8390}{0.0736}\right)} = \frac{0.0144}{0.9856} \simeq 0.0146.$$

Quindi $BF_{01} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{0.0146}{24.4453} = 0.0005972518$, e $-2\log(BF_{01}) \simeq -14.84$. Il dati mostrano una fortissima evidenza a favore di H_1.

∎

STATISTICA BAYESIANA A. Guglielmi e R. Argiento 14.07.2015

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 Si consideri il seguente modello:

$$Y_1, \dots, Y_{M_1} | \lambda_1, M_1 \overset{\text{iid}}{\sim} Poi(\lambda_1), \quad Y_{M_1+1}, \dots, Y_n | \lambda_2, M_1 \overset{\text{iid}}{\sim} Poi(\lambda_2), \quad \lambda_1, \lambda_2 > 0, \qquad (1)$$

con $M_1 \in \{1, 2, \dots, n-1\}$ e i due blocchi di variabili sono condizionatamente indipendenti. Si indichi con $M_2 = n - M_1$ la dimensione del secondo blocco di osservazioni.

Per il momento, si supponga che il *change point* M_1 sia noto e pari a $m_1 \in \{1, 2, \dots, n-1\}$.

1. Determinare la prior $\pi_1(\lambda_1) \times \pi_2(\lambda_2)$ coniugata, scrivendo esplicitamente l'aggiornamento dei parametri. Per unificare la notazione, indicare con (α_j, β_j) gli iperparametri della prior marginale $\pi_j(\lambda_j)$, per $j = 1, 2$ e con α_{1m_1}, β_{1m_1}, α_{2m_2}, β_{2m_2} gli iperparametri aggiornati.

2. Per fissare gli iperparametri della prior coniugata ricavata al punto 1., un esperto ha fissato i valori marginali a priori $\mathrm{E}(Y_1) = 3$, $\mathrm{Var}(Y_1) = 6$, $\mathrm{E}(Y_n) = 6$, $\mathrm{Var}(Y_n) = 8$. Ricavare i valori di α_1, β_1, α_2, β_2, sulla base di tali informazioni.

Ora invece si assuma che M_1 sia incognito, e che si voglia fare inferenza su di esso, sulla base della prior uniforme su $\{1, 2, \dots, n-1\}$; si supponga inoltre che M_1 e (λ_1, λ_2) siano a priori indipendenti e (λ_1, λ_2) abbia distribuzione a priori come ricavata ai punti 1. e 2.

3. Determinare la forma analitica della distribuzione finale di M_1, anche a meno della costante di normalizzazione.
   Suggerimento: Si consiglia dapprima di ricavare la distribuzione finale congiunta di $(M_1, \lambda_1, \lambda_2)$ e successivamente ricavare la distribuzione marginale a posteriori di M_1.

4. I dati a disposizione sono descritti nella seguente tabella:

Tabella 1: Dati osservati.

i	1	2	3	4	5	6	7	8	9	10	11
y_i	2	4	2	5	6	0	6	9	6	6	10
s_i	2	6	8	13	19	19	25	34	40	46	56
$s_n - s_i$	54	50	48	43	37	37	31	22	16	10	
$\log\Gamma(3 + s_i)$	3.18	10.60	15.10	27.90	45.38	45.38	64.56	95.72	117.77	140.67	
$\log\Gamma(18 + s_n - s_i)$	234.70	217.74	209.34	188.63	164.32	164.32	140.67	106.63	85.05	64.56	

   dove $s_i := \sum_{j=1}^{i} y_j$ e $s_n - s_i := \sum_{j=i+1}^{n} y_j$. Sulla base della posterior al punto 3, con i dati in tabella, ricavare la moda a posteriori di M_1 per stimare il *change point*.

5. Si vogliono confrontare le ipotesi H_1: $M_1 = m_1$ vs H_2: $M_1 = m_1+1$, dove m_1 è un intero fissato in $\{1, 2, \dots, n-2\}$. Anzitutto, ricavare l'espressione analitica del Bayes factor BF_{12} di H_1 contro H_2 in funzione degli iperparametri a priori e del dataset $(y_1, \dots, y_{m_1}, y_{m_1+1}, \dots, y_n)$. Successivamente mostrare che se $y_{m_1+1} = 0$, allora l'espressione di BF_{12} è funzione dei soli iperparametri a posteriori α_{1m_1}, β_{1m_1}, α_{2m_2}, β_{2m_2}; determinare esplicitamente tale espressione.

6. Sfruttando i risultati al punto 5., sulla base dei dati in tabella, quale ipotesi si sceglie tra H_1: $M_1 = 5$ vs H_2: $M_1 = 6$? Giustificare la risposta.

Soluzione

1. Siccome la verosimiglianza $\mathcal{L}(\lambda_1, \lambda_2; m_1, y_1, \ldots, y_{m_1}, y_{m_1+1}, \ldots, y_n) = \mathcal{L}(\lambda_1; m_1, y_1, \ldots, y_{m_1}) \times \mathcal{L}(\lambda_2; m_1, y_{m_1+1}, \ldots, y_n)$ e $\pi(\lambda_1, \lambda_2) = \pi_1(\lambda_1) \times \pi_2(\lambda_2)$, anche a posteriori λ_1 e λ_2 sono indipendenti; sappiamo, inoltre, che la prior coniugata al modello Poisson è la gamma. Quindi:

$$\begin{aligned}
&\pi(\lambda_1, \lambda_2 | m_1, y_1, \ldots, y_{m_1}, y_{m_1+1}, \ldots, y_n) = \pi_1(\lambda_1 | m_1, y_1, \ldots, y_{m_1}) \times \pi_2(\lambda_2 | m_1, y_{m_1+1}, \ldots, y_n) \\
&= \prod_{i=1}^{m_1} \mathrm{e}^{-\lambda_1} \frac{\lambda_1^{y_i}}{y_i!} \prod_{i=m_1+1}^{n} \mathrm{e}^{-\lambda_2} \frac{\lambda_2^{y_i}}{y_i!} \times \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \lambda_1^{\alpha_1 - 1} \mathrm{e}^{-\beta_1 \lambda_1} \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \lambda_2^{\alpha_2 - 1} \mathrm{e}^{-\beta_2 \lambda_2} \mathbf{1}_{(0,+\infty)}(\lambda_1) \mathbf{1}_{(0,+\infty)}(\lambda_2) \\
&\propto \lambda_1^{\alpha_1 + \sum_1^{m_1} y_i - 1} \mathrm{e}^{-(\beta_1 + m_1)\lambda_1} \mathbf{1}_{(0,+\infty)}(\lambda_1) \times \lambda_2^{\alpha_2 + \sum_{m_1+1}^{n} y_i - 1} \mathrm{e}^{-(\beta_2 + n - m_1)\lambda_2} \mathbf{1}_{(0,+\infty)}(\lambda_2) \\
&= \mathrm{gamma}(\alpha_{1m_1}, \beta_{1m_1}) \times \mathrm{gamma}(\alpha_{2m_2}, \beta_{2m_2}),
\end{aligned}$$

$\alpha_{1m_1} = \alpha_1 + \sum_1^{m_1} y_i$, $\beta_{1m_1} = \beta_1 + m_1$, $\alpha_{2m_2} = \alpha_2 + \sum_{m_1+1}^{n} y_i$, $\beta_{2m_2} = \beta_2 + n - m_1 = \beta_2 + m_2$.

2. Si ricava:

$$\begin{aligned}
\mathrm{E}(Y_1) &= \mathrm{E}(\mathrm{E}(Y_1 | \lambda_1)) = \mathrm{E}(\lambda_1) = \frac{\alpha_1}{\beta_1} \\
\mathrm{Var}(Y_1) &= \mathrm{E}(\mathrm{Var}(Y_1 | \lambda_1)) + \mathrm{Var}(\mathrm{E}(Y_1 | \lambda_1)) = \mathrm{E}(\lambda_1) + \mathrm{Var}(\lambda_1) = \frac{\alpha_1}{\beta_1} + \frac{\alpha_1}{\beta_1^2} = \frac{\alpha_1(\beta_1 + 1)}{\beta_1^2}.
\end{aligned}$$

Si tratta quindi di risolvere il sistema di equazioni:

$$\frac{\alpha_1}{\beta_1} = 3, \quad \frac{\alpha_1(\beta_1 + 1)}{\beta_1^2} = 6,$$

la cui soluzione è $\alpha_1 = 3$, $\beta_1 = 1$.

In modo analogo si trova $\mathrm{E}(Y_n) = \alpha_2/\beta_2$ e $\mathrm{Var}(Y_n) = \alpha_2(\beta_2 + 1)/\beta_2^2$, e il secondo sistema di equazioni è

$$\frac{\alpha_2}{\beta_2} = 6, \quad \frac{\alpha_2(\beta_2 + 1)}{\beta_2^2} = 8,$$

la cui soluzione è $\alpha_2 = 18$, $\beta_2 = 3$.

3. Ora anche M_1 è incognito. Pertanto:

$$\begin{aligned}
&\pi(m_1, \lambda_1, \lambda_2 | y_1, \ldots, y_{m_1}, y_{m_1+1}, \ldots, y_n) \\
&\propto \lambda_1^{\alpha_1 + \sum_1^{m_1} y_i - 1} \mathrm{e}^{-(\beta_1 + m_1)\lambda_1} \lambda_2^{\alpha_2 + \sum_{m_1+1}^{n} y_i - 1} \mathrm{e}^{-(\beta_2 + m_2)\lambda_2} \frac{1}{n-1} \mathbf{1}_{\{1,2,\ldots,n-1\}}(m_1) \mathbf{1}_{(0,+\infty)}(\lambda_1) \mathbf{1}_{(0,+\infty)}(\lambda_2).
\end{aligned}$$

Allora, per $m_1 = 1, 2, \ldots, n-1$, si ha

$$\begin{aligned}
\mathbb{P}(M_1 = m_1 | dati) &= \int_0^{+\infty} \int_0^{+\infty} \pi(m_1, \lambda_1, \lambda_2 | m_1, y_1, \ldots, y_{m_1}, y_{m_1+1}, \ldots, y_n) d\lambda_1 d\lambda_2 \\
&\propto \frac{\Gamma(\alpha_1 + \sum_1^{m_1} y_j)}{(\beta_1 + m_1)^{\alpha_1 + \sum_1^{m_1} y_j}} \frac{\Gamma(\alpha_2 + \sum_{m_1+1}^{n} y_j)}{(\beta_2 + n - m_1)^{\alpha_2 + \sum_{m_1+1}^{n} y_j}} \mathbf{1}_{\{1,2,\ldots,n-1\}}(m_1). \qquad (2)
\end{aligned}$$

4. Il punto di massimo di $\mathbb{P}(M_1 = m_1 | dati)$ (la moda a posteriori) si trova massimizzando, rispetto ad $m_1 = 1, 2, \ldots, n-1$, l'espressione (2), con i dati in Tabella 1. Calcoliamo il logaritmo dei valori in (2); per esempio,
se $i = 1$, $\log\Gamma(3 + s_1) + \log\Gamma(18 + s_{11} - s_1) - (3 + s_1)\log 2 - (18 + s_{11} - s_1)\log 13 = 49.73791 \simeq 49.74$. In modo analogo, si ottengono tutti gli altri valori:

i	1	2	3	4	5	6	7	8	9	10
log di (2)	49.74	49.48	50.93	50.32	49.43	52.52	51.66	49.38	49.09	48.92

Dunque il massimo si ottiene per $m_1 = 6$ e $\widehat{M_1} = 6$, moda a posteriori di M_1, è la stima di M_1.

5. Il Bayes factor per le ipotesi H_1: $M_1 = m_1$ vs H_2: $M_1 = m_1+1$, dove m_1 è un intero fissato in $\{1, 2, \ldots, n-1\}$, coincide con il posterior odd, visto che la prior per M_1 è uniforme; pertanto:

$$BF_{12} = \frac{\mathbb{P}(M_1 = m_1|dati)}{\mathbb{P}(M_1 = m_1 + 1|dati)}$$
$$= \frac{\frac{\Gamma(\alpha_1+\sum_1^{m_1} y_j)}{(\beta_1+m_1)^{\alpha_1+\sum_1^{m_1} y_j}} \frac{\Gamma(\alpha_2+\sum_{m_1+1}^{n} y_j)}{(\beta_2+n-m_1)^{\alpha_2+\sum_{m_1+1}^{n} y_j}}}{\frac{\Gamma(\alpha_1+\sum_1^{m_1} y_j+y_{m_1+1})}{(\beta_1+m_1+1)^{\alpha_1+\sum_1^{m_1} y_j+y_{m_1+1}}} \frac{\Gamma(\alpha_2+\sum_{m_1+2}^{n} y_j)}{(\beta_2+n-m_1-1)^{\alpha_2+\sum_{m_1+2}^{n} y_j}}}.$$

Ora, se $y_{m_1+1} = 0$, si ha

$$BF_{12} = \frac{\frac{\Gamma(\alpha_1+\sum_1^{m_1} y_j)}{(\beta_1+m_1)^{\alpha_1+\sum_1^{m_1} y_j}} \frac{\Gamma(\alpha_2+\sum_{m_1+1}^{n} y_j)}{(\beta_2+n-m_1)^{\alpha_2+\sum_{m_1+1}^{n} y_j}}}{\frac{\Gamma(\alpha_1+\sum_1^{m_1} y_j)}{(\beta_1+m_1+1)^{\alpha_1+\sum_1^{m_1} y_j}} \frac{\Gamma(\alpha_2+\sum_{m_1+1}^{n} y_j)}{(\beta_2+n-m_1-1)^{\alpha_2+\sum_{m_1+1}^{n} y_j}}}$$
$$= \left(\frac{\beta_1 + m_1 + 1}{\beta_1 + m_1}\right)^{\alpha_1+\sum_1^{m_1} y_j} \left(\frac{\beta_2 + n - m_1 - 1}{\beta_2 + n - m_1}\right)^{\alpha_2+\sum_{m_1+1}^{n} y_j}$$
$$= \left(\frac{\beta_{1,m_1} + 1}{\beta_{1,m_1}}\right)^{\alpha_{1,m_1}} \left(\frac{\beta_{2,m_2} - 1}{\beta_{2,m_2}}\right)^{\alpha_{2,m_2}}.$$

6. Si applica il risultato trovato sopra al caso $m_1 = 5$.
Dunque, se H_1: $M_1 = 5$ vs H_2: $M_1 = 6$, si ha:

$$BF_{12} = \left(\frac{\beta_{1,5} + 1}{\beta_{1,5}}\right)^{\alpha_{1,5}} \left(\frac{\beta_{2,6} - 1}{\beta_{2,6}}\right)^{\alpha_{2,6}},$$

dove $m_1 = 5$, $m_2 = n - m_1 = 6$, $\alpha_{1,5} = \alpha_1 + s_5 = 3 + 19 = 22$, $\beta_{1,5} = \beta_1 + 5 = 6$, $\alpha_{2,6} = \alpha_2 + s_{11} - s_5 = 18 + 37 = 55$, $\beta_{2,6} = \beta_2 + 6 = 9$; quindi

$$BF_{12} = \left(\frac{7}{6}\right)^{22} \left(\frac{8}{9}\right)^{55} = 0.04565,$$

e $2 \log BF_{12} \simeq -6.17$, indicando forte evidenza sperimentale a favore di H_2: $M_1 = 6$.

∎

STATISTICA BAYESIANA A. Guglielmi e R. Argiento 26.02.2015

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 Si vuole confrontare l'efficacia di due farmaci. A tale scopo si testano in modo indipendente i due medicinali su due campioni di pazienti di ampiezza n_1 e n_2, rispettivamente. Il campione a disposizione è rappresentato da (X_1, X_2), dove X_i è il numero di pazienti del gruppo i sui quali il farmaco i è risultato efficace, con $i = 1, 2$. Si assume che, condizionatamente a (θ_1, θ_2), con $\theta_1, \theta_2 \in (0, 1)$, X_1 e X_2 sono indipendenti e

$$X_i \sim \text{Binomiale}(n_i, \theta_i) \quad i = 1, 2.$$

Si vuole fare inferenza su θ_1, θ_2, e $\varphi = \log(\theta_1/\theta_2)$. Quest'ultima variabile è usata dagli epidemiologi per riassumere i risultati dell'esperimento. Si osservi che se $\varphi < 0$ ($\varphi > 0$), il farmaco 1 è meno (più) efficace del farmaco 2.

1. Determinare la prior $\pi_1(\theta_1) \times \pi_2(\theta_2)$ coniugata e scrivere esplicitamente l'aggiornamento dei parametri. Per unificare la notazione si indichino con (α_i, β_i) gli iperparametri della prior marginale $\pi_i(\theta_i)$, per $i = 1, 2$.

2. I farmaci hanno già passato dei test. Entrambi sono stati somministrati a due campioni di ampiezza 10, e sono risultati efficaci su 7 pazienti. Utilizzare questa informazione per fissare gli iperparametri della prior coniugata ricavata al punto 1. utilizzando il principio del campione equivalente.

3. Sapendo ora che $n_1 = 35$, $x_1 = 29$ e $n_2 = 32$, $x_2 = 24$, ricavare il valore degli iperparametri della posterior di (θ_1, θ_2) determinata ai punti precedenti.

4. È necessario ora fare dei conti che serviranno per i punti successivi. Se X è una variabile aleatoria beta(α, β), ricavare $\mathrm{E}((1-X)^k)$ per $k \geq 1$, k intero.

5. Calcolare media e varianza a posteriori del parametro φ. A tal fine, si utilizzeranno opportune approssimazioni (dallo sviluppo in serie di Taylor delle funzioni $\log x$ e $(\log x)^2$) per il calcolo dei momenti primo e secondo a posteriori di $\log(\theta_i)$, quando θ_i ha una opportuna distribuzione beta. In particolare, per $i = 1, 2$, utilizzare

$$\mathrm{E}(\log(\theta_i)|n_i, x_i) \simeq -\sum_{l=1}^{5} \frac{1}{l}\,\mathrm{E}((1-\theta_i)^l)|n_i, x_i) \tag{1}$$

$$\begin{aligned}\mathrm{E}(\log^2(\theta_i)|n_i, x_i) \simeq{}& \mathrm{E}((1-\theta_i)^2|n_i, x_i) + \mathrm{E}((1-\theta_i)^3|n_i, x_i)\\ &+ \frac{11}{12}\,\mathrm{E}((1-\theta_i)^4|n_i, x_i) + \frac{5}{6}\,\mathrm{E}((1-\theta_i)^5|n_i, x_i).\end{aligned} \tag{2}$$

6. Ricorrendo ad una approssimazione gaussiana per la distribuzione finale di $\log(\theta_i)$, per $i = 1, 2$, calcolare la probabilità a posteriori che il farmaco 1 sia meno efficace del farmaco 2.

Esercizio 2 Ricavare la legge marginale di $(X_1, \ldots, X_n)$, campione aleatorio da un processo di Dirichlet P di parametri $\alpha > 0$ e P_0 (misura di probabilità su $\mathbb{R}$).

Soluzione

1. Siccome la verosimiglianza $\mathcal{L}(\theta_1, \theta_2; n_1, x_1, n_2, x_2) = \mathcal{L}(\theta_1; n_1, x_1) \times \mathcal{L}(\theta_2; n_2, x_2)$ e $\pi(\theta_1, \theta_2) = \pi_1(\theta_1) \times \pi_2(\theta_2)$, anche a posteriori θ_1 e θ_2 sono indipendenti, cioè

$$\pi(\theta_1, \theta_2 | n_1, x_1, n_2, x_2) = \pi_1(\theta_1 | n_1, x_1) \times \pi_2(\theta_2 | n_2, x_2).$$

D'altro canto, sappiamo che la prior coniugata al modello binomiale è la beta. Pertanto $\pi(\theta_1) \times \pi(\theta_2) =$beta$(\alpha_1, \beta_1)\times$beta$(\alpha_2, \beta_2)$. Dunque, a posteriori θ_1 e θ_2 sono indipendenti, con distribuzione $beta(\widetilde{\alpha}_1, \widetilde{\beta}_1) \times beta(\widetilde{\alpha}_2, \widetilde{\beta}_2)$, con
$\widetilde{\alpha}_1 = \alpha_1 + x_1$; $\widetilde{\beta}_1 = \beta_1 + n_1 - x_1$; $\widetilde{\alpha}_2 = \alpha_2 + x_2$; $\widetilde{\beta}_2 = \beta_2 + n_2 - x_2$.

2. Scriviamo la media a posteriori di θ_i per $i = 1, 2$:

$$\mathrm{E}(\theta_i | n_i, x_i) = \frac{\alpha_i + x_i}{\alpha_i + \beta_i + n_i} = \frac{\alpha_i + \beta_i}{\alpha_i + \beta_i + n_i} \frac{\alpha_i}{\alpha_i + \beta_i} + \frac{n_i}{\alpha_i + \beta_i + n_i} \frac{x_i}{n_i}$$

Si vede che $\alpha_i/(\alpha_i + \beta_i)$ ha il significato del numero di successi diviso il numero delle prove (in una successione di prove di Bernoulli) e cioè α_i ha lo stesso significato del numero di successi e $\alpha_i + \beta_i$ ha lo stesso significato del numero di prove nel campione equivalente. Quindi $\alpha_i = 7$ e $\beta_i = 10 - 7 = 3$, per $i = 1, 2$.

3. Per quanto riguarda gli aggiornamenti degli iperparametri della coniugata, si ha:
$\widetilde{\alpha}_1 = \alpha_1 + x_1 = 7 + 29 = 36$;
$\widetilde{\beta}_1 = \beta_1 + n_1 - x_1 = 3 + 35 - 29 = 9$;
$\widetilde{\alpha}_2 = \alpha_2 + x_2 = 7 + 24 = 31$;
$\widetilde{\beta}_2 = \beta_2 + n_2 - x_2 = 3 + 32 - 24 = 11$.

4. Per $k = 1, 2, 3 \ldots$, si ricava:

$$\begin{aligned}
\mathrm{E}\left((1 - X)^k\right) &= \int_0^1 (1 - x)^k \frac{1}{B(\alpha, \beta)} x^{\alpha - 1}(1 - x)^{\beta - 1} dx = \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha - 1}(1 - x)^{k + \beta - 1} dx \\
&= \frac{B(\alpha, \beta + k)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha)\Gamma(\beta + k)}{\Gamma(\alpha + \beta + k)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(\beta + k)}{\Gamma(\beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + k)} \\
&= \frac{(\beta + k - 1) \cdots (\beta + 1)\beta}{(\alpha + \beta + k - 1) \cdots (\alpha + \beta + 1)(\alpha + \beta)}.
\end{aligned}$$

Tra l'altro, si noti che per $k = 1$

$$\mathrm{E}(1 - X) = \frac{\Gamma(\beta + 1)}{\Gamma(\alpha + \beta + 1)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)} = \frac{\beta\Gamma(\beta)}{(\alpha + \beta)\Gamma(\alpha + \beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)} = \frac{\beta}{\alpha + \beta}$$

e che per $k \geq 2$

$$\mathrm{E}\left((1 - X)^k\right) = \frac{\beta + k - 1}{\alpha + \beta + k - 1} \mathrm{E}\left((1 - X)^{k-1}\right).$$

5. Osserviamo innanzitutto che

$$\mathrm{E}(\varphi | x_1, x_2) = \mathrm{E}(\log(\theta_1) | x_1) - \mathrm{E}(\log(\theta_2 | x_2)).$$

Utilizzando i valori dei parametri a posteriori si ha che i primi cinque momenti a posteriori di 1-θ_1 valgono (nell'ordine) 0.2000, 0.0435, 0.0102, 0.0025, 0.0007, mentre i primi cinque momenti a posteriori di 1-θ_2 valgono (nell'ordine) 0.2619, 0.0731, 0.0216, 0.0067, 0.0022.

Utilizzando la (1), si trova che

$$\begin{aligned}
\mathrm{E}(\log(\theta_1)|n_1, x_1) &\simeq -0.2000 - \frac{1}{2} \times 0.0435 - \frac{1}{3} \times 0.0102 - \frac{1}{4} \times 0.0025 - \frac{1}{5} \times 0.0007 = -0.225915 \\
&\simeq -0.2260 \\
\mathrm{E}(\log(\theta_2)|n_2, x_2) &\simeq -0.2619 - \frac{1}{2} \times 0.0731 - \frac{1}{3} \times 0.0216 - \frac{1}{4} \times 0.0067 - \frac{1}{5} \times 0.0022 = -0.307765 \\
&\simeq -0.3078,
\end{aligned}$$

e dunque si ricava $\mathrm{E}(\varphi|x_1, x_2) \simeq -0.226 + 0.3078 = 0.0818$.

Vogliamo ora valutare

$$\mathrm{E}(\varphi^2|n_1, x_1, n_2, x_2) = \mathrm{E}(\log^2(\theta_1)|n_1, x_1) + \mathrm{E}(\log^2(\theta_2)|n_2, x_2) - 2\,\mathrm{E}(\log(\theta_1)|n_1, x_1)\,\mathrm{E}(\log(\theta_2)|n_2, x_2).$$

Per quanto riguarda i momenti secondi presenti nella formula qui sopra, utilizzando l'approssimazione nella (2), si ottiene

$$\begin{aligned}
\mathrm{E}(\log^2(\theta_1)|n_1, x_1) &\simeq 0.0435 + 0.0102 + \frac{11}{12}0.0025 + \frac{5}{6}0.0007 = 0.056575 \simeq 0.0566 \\
\mathrm{E}(\log^2(\theta_2)|n_2, x_2) &\simeq 0.0731 + 0.0216 + \frac{11}{12}0.0067 + \frac{5}{6}0.0022 = 0.102675 \simeq 0.1027\ .
\end{aligned}$$

In conclusione si ricava che

$$\mathrm{E}(\varphi^2|n_1, x_1, n_2, x_2) = 0.0566 + 0.103 - 2 \times 0.226 \times 0.308 = 0.0201744 \simeq 0.0202,$$

e $\mathrm{Var}(\varphi|n_1, x_1, n_2, x_2) = \mathrm{E}(\varphi^2|n_1, x_1, n_2, x_2) - \mathrm{E}(\varphi|n_1, x_1, n_2, x_2)^2 = 0.0135$.

6. Con un certo grado di approssimazione, $\log\theta_1$ e $\log\theta_2$, a posteriori, hanno distribuzioni marginali gaussiane (e sono indipendenti). Allora φ a posteriori ha approssimativamente distribuzione $\mathcal{N}(\mathrm{E}(\varphi), \mathrm{Var}(\varphi))$=$\mathcal{N}(0.0818, 0.0135)$. Quindi si deve calcolare

$$\mathbb{P}(\varphi < 0|n_1, x_1, n_2, x_2) = \mathbb{P}\left(\varphi < -\frac{0.0818}{\sqrt{0.0135}}|n_1, x_1, n_2, x_2\right) \simeq \mathbb{P}(Z < -0.7040) = 0.2407$$

dove $Z \sim$N$(0, 1)$.

$\blacksquare$

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 Sia $\underline{X} = (X_1, \ldots, X_n)$ un campione di dati a valori in $\mathbb{R}^+$. Condizionatamente ai parametri $\theta > 0$ (scale) e $\beta > 0$ (shape), assumiamo che i dati X_i siano indipendenti e identicamente distribuiti con distribuzione Weibull(θ, β), cioè

$$f_{X_i}(x; \theta, \beta) = \frac{\beta}{\theta} x^{\beta-1} \exp\left\{-\frac{x^\beta}{\theta}\right\} \mathbf{1}_{(0,+\infty)}(x). \tag{1}$$

1. Supponendo che β sia noto, ricavare la prior coniugata $\pi(\theta)$ per θ. Scrivere esplicitamente l'aggiornamento dei parametri della posterior, utilizzando la parametrizzazione di $\pi(\theta)$ tale che $E_\pi(\theta) = b/(a-1)$ (quando la media esiste).
   Suggerimento: potrebbe essere utile la notazione $s(\beta, \underline{x}) = \sum_1^n x_i^\beta$.

Si consideri ora un campione $\underline{z} = (z_1, \ldots, z_m)$ di dimensione m, ottenuto sotto le stesse condizioni di $\underline{x}$. Si vuole utilizzare tale campione per fissare gli iperparametri della prior coniugata trovata al punto 1. Si adottano due strategie, descritte ai punti 2 e 3 successivi; in entrambi i casi gli iperparametri della prior dipenderanno dal parametro noto β, oltre che da $\underline{z}$.

2. **Campione equivalente**: utilizzare il principio del campione equivalente per fissare gli iperparametri a_1 e b_1 della prior coniugata $\pi_1(\theta) := \pi(\theta; a_1, b_1)$.

3. **Metodo marginale**: si sceglie la prior coniugata $\pi_2(\theta) := \pi(\theta; a_2, b_2)$ con $a_2 > 0$ fissato e $b_2 = \text{argmax}_b(m(\underline{z}); a_2, b)$, dove $m(\underline{z}; a, b)$ è la marginale del campione equivalente $\underline{z}$ quando la prior è quella coniugata $\pi(\theta; a, b)$.

Mediante l'utilizzo del Bayes factor, sulla base dei dati $(x_1, \ldots, x_n)$, si vogliono confrontare due modelli: la verosimiglianza è la medesima, ed è quella specificata da (1), mentre la prior è $\pi_1(\theta)$ per il modello 1 e $\pi_2(\theta)$ per il modello 2.

4. Ricavare l'espressione del Bayes factor tra i modelli 1 e 2. In particolare, scrivere l'espressione del Bayes factor quando $a_2 = m$.

5. Si supponga ora che $\beta = 2$. Il campione storico è $\underline{z} = (1.51, 0.50, 1.2, 0.32)$ mentre il campione osservato è $\underline{x} = (0.54, 0.69, 5.47, 5.29, 4.13, 4.51, 3.68, 4.96, 3.70, 1.54)$. Utilizzando il modello migliore, rispetto al Bayes factor del punto 4 (ponendo $a_2 = m$), fornire la stima bayesiana a posteriori di θ rispetto alla funzione di perdita quadratica.

Si vuole ora fare inferenza anche sul parametro β, sempre nel caso di dati $\underline{X} = (X_1, \ldots, X_n)$ condizionatamente i.i.d. dalla densità (1). Si assume che a priori θ e β in (1) siano indipendenti, β con distribuzione $gamma(c, d)$, mentre per θ si assume la prior coniugata $\pi(\theta; a, b)$ trovata al punto 1. I parametri a, b, c, d sono tutti positivi.

6. Scrivere le due full-conditionals per implementare un algoritmo di tipo Gibbs sampler per il calcolo della posterior di (θ, β), con i dati $\underline{x} = (x_1, \ldots, x_n)$.

7. Assumendo $d > \sum_1^n \log(x_i)$, mostrare che è possibile utilizzare un metodo acceptance-rejection con proposal gamma per campionare dalla full-conditional di β. Di conseguenza, determinare esplicitamente una coppia di parametri per la proposal gamma.

Soluzione

1. La verosimiglianza condizionale del modello Weibull è

$$L(\theta,\beta;\underline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i;\alpha,\beta) = \frac{\beta^n}{\theta^n}\left(\prod_{i=1}^{n} x_i\right)^{\beta-1} \exp\left\{-\frac{\sum_{i=1}^{n} x_i^{\beta}}{\theta}\right\},\ \theta,\beta>0$$

dove $s(\beta,\underline{x}) := \sum_{i=1}^{n} x_i^{\beta}$. Visto che β è noto,

$$L(\theta;\underline{x},\beta) \propto \frac{1}{\theta^n}\exp\left\{-\frac{s(\beta,\underline{x})}{\theta}\right\}\mathbf{1}_{(0,+\infty)}(\theta),$$

che è il kernel di una densità inversa-gamma. Consideriamo quindi come prior

$$\pi(\theta;a,b) = \frac{b^a}{\Gamma(a)}\frac{1}{\theta^{a+1}}\exp\left\{-\frac{b}{\theta}\right\}\mathbf{1}_{0,+\infty}(\theta)$$

dove a è il parametro shape e b è il rate. In tal caso $\mathrm{E}(\theta) = b/(a-1)$ se $a>1$. Utilizzando il teorema di Bayes

$$\pi(\theta|\underline{x}) \propto \frac{1}{\theta^{n+a+1}}\exp\left\{-\frac{s(\beta,\underline{x})+b}{\theta}\right\}\mathbf{1}_{(0,+\infty)}(\theta)$$

A posteriori, θ è distribuito come una inversa gamma di parametri $a_n = a+n$ e $b_n = b+s(\beta,\underline{x})$.

2. Come risultato preliminare notiamo che lo stimatore di massima verosimiglianza per θ è $\hat{\theta}(\beta) = \frac{s(\beta,\underline{x})}{n}$. Infatti

$$\log(L(\theta,\beta;\underline{x}) = n\log(\beta) - n\log(\theta) + (\beta-1)\sum_{1}^{n}\log(x_i) - \frac{s(\beta,\underline{x})}{n}.$$

Quindi $\frac{d}{d\theta}\log(L(\theta,\beta;\underline{x}) > 0$ se e solo se $\theta < \frac{s(\beta,\underline{x})}{n}$.

Osserviamo ora che la media a posteriori di θ è:

$$\mathrm{E}(\theta|\underline{x}) = \frac{b+s(\beta,\underline{x})}{a+n-1} = \frac{a-1}{n+a-1}\frac{b}{a-1}+\frac{n}{a+n-1}\frac{s(\beta,\underline{x})}{n} = \frac{a-1}{n+a-1}\mathrm{E}(\theta)+\frac{n}{a+n-1}\hat{\theta}(\beta).$$

Naturalmente, che questo ultimo passaggio ha senso solo se $a>1$. Utilizzando il principio del campione equivalente si ha che $a_1 = m+1$ e $b_1 = s(\beta,\underline{z})$.

3. Innanzitutto calcoliamo la marginale del campione $\underline{z} = (z_1,\ldots,z_m)$:

$$\begin{aligned}
m(\underline{z};a,b) &= \int_0^{+\infty}\prod_{i=1}^{m} f_{X_i}(z_i;\beta,\theta)\pi(\theta;a,b)d\theta \\
&= \int_0^{\infty}\frac{\beta^m}{\theta^m}\left(\prod_{i=1}^{m} z_i\right)^{\beta-1}\exp\left\{-\frac{\sum_{i=1}^{m} z_i^{\beta}}{\theta}\right\}\frac{b^a}{\Gamma(a)}\frac{1}{\theta^{a+1}}\exp\left\{-\frac{b}{\theta}\right\}d\theta \\
&= \beta^m\left(\prod_{1}^{m} z_i\right)^{\beta-1}\frac{\Gamma(a+m)}{\Gamma(a)}\frac{b^a}{\left(b+\sum_{i=1}^{m} z_i^{\beta}\right)^{m+a}}
\end{aligned}$$

Si ha quindi che

$$\frac{d}{db}m(\underline{z};a_2,b) \propto \frac{d}{db}\frac{b_2^a}{(b+s(\beta,\underline{z})))^{m+a_2}}$$
$$= \frac{a_2 b^{a_2-1}(b+s(\beta,\underline{z}))^{a_2+m} - b^{a_2}(a_2+m)(s(\beta,\underline{z})+b)^{a_2+m-1}}{(s(\beta,\underline{z})+b)^{2(m+a_2)}}$$

Pertanto $\frac{d}{db}m(\underline{z};a,b) \geq 0$ se e solo se $mb - a_2 s(\beta,\underline{z}) \leq 0$. In conclusione $b_2 = a_2\frac{s(\beta,\underline{z})}{m}$.

4. Il Bayes factor è:

$$BF_{12} = \frac{m(\underline{x};a_1,b_1)}{m(\underline{x};a_2,b_2)} = \frac{\beta^n\left(\prod_1^n x_i\right)^{\beta-1}\frac{\Gamma(a_1+n)}{\Gamma(a_1)}\frac{b_1^{a_1}}{(b_1+\sum_i^n x_i^\beta)^{a_1+n}}}{\beta^n\left(\prod_1^n x_i\right)^{\beta-1}\frac{\Gamma(a_2+n)}{\Gamma(a_2)}\frac{b_2^{a_2}}{(b_2+\sum_i^n x_i^\beta)^{a_2+n}}}$$
$$= \frac{\Gamma(a_2)\,\Gamma(a_1+n)\,b_1^{a_1}}{\Gamma(a_1)\,\Gamma(a_2+n)\,b_2^{a_2}}\frac{\left(b_2+\sum_1^n x_i^\beta\right)^{a_2+n}}{\left(b_1+\sum_1^n x_i^\beta\right)^{a_1+n}}$$

Ora, se $a_2 = m$, si ricava $b_2 = s(\beta,\underline{z}) = \sum_1^m z_i^\beta$ e si ha:

$$BF_{12} = \frac{\Gamma(m)}{\Gamma(m+1)}\frac{\Gamma(m+n+1)}{\Gamma(m+n)}\frac{s(\beta,\underline{z})^{m+1}}{s(\beta,\underline{z})^m}\frac{1}{s(\beta,\underline{z})+\sum_1^n x_i^\beta}$$
$$= \frac{m+n}{m}\frac{\sum_1^m z_i^\beta}{\sum_1^m z_i^\beta+\sum_1^n x_i^\beta}.$$

5. Sostituendo i valori osservati nella formula ricavata al punto precedente, otteniamo

$$BF_{12} = \frac{10+4}{4}\frac{4.0725}{4.0725+150.2753} = 0.0923.$$

Quindi $2\log(BF_{12}) = -4.7654$ e secondo la scala di Jeffreys c'è evidenza empirica a favore del modello 2.

Fissando gli iperparametri della prior per θ secondo il modello 2, si trova $a_2 = m = 4$ e $b_2 = s(2,\underline{z}) = 4.0725$. La stima bayesiana (a posteriori) del parametro θ è la media a posteriori, in questo caso. Quindi

$$\mathrm{E}(\theta|\underline{x}) = \frac{b_2+s(2,\underline{x})}{a_2+n-1} = 11.8729.$$

6. La full conditional di θ coincide con la posterior trovata al punto 1, e cioè una inversa gamma di parametri $a_n = a+n$ e $b_n = b+\sum_1^n x_i^\beta$.

Per quanto riguarda β si ha che

$$\pi(\beta|\theta,\underline{x}) \propto \beta^n\left(\prod_{i=1}^n x_i\right)^{\beta-1}\exp\left\{-\frac{\sum_{i=1}^n x_i^\beta}{\theta}\right\}\beta^{c-1}\exp(-d\beta)\mathbf{1}_{(0,+\infty)}(\beta)$$
$$= \beta^{n+c-1}\exp\left\{-\beta\left(d-\sum_{i=1}^n\log(x_i)\right)\right\}\exp\left\{-\frac{s(\beta,\underline{x})}{\theta}\right\}\mathbf{1}_{(0,+\infty)}(\beta)$$

Quest'ultima espressione non coincide con alcun kernel di densità notevoli.

7. Ricordiamo che l'algoritmo di accettazione-rifiuto può essere applicato anche se la distribuzione target è nota a meno della costante di normalizzazione, come nel nostro caso.

   Poniamo $f(\beta) := \beta^{n+c-1} \exp\{-\beta(d - \sum_{i=1}^n \log(x_i))\} \exp\left\{-\frac{s(\beta,\underline{x})}{\theta}\right\} \mathbf{1}_{(0,+\infty)}(\beta)$. Bisogna mostrare che

$$f(\beta) \leq Mg(\beta), \quad M > 0,$$

   dove g è il kernel di una distribuzione gamma.

   Osserviamo ora che $\exp\left\{-\frac{s(\beta,\underline{x})}{\theta}\right\} < 1$ per ogni $\beta > 0$; quindi si ha che

$$g(\beta) < \beta^{n+c-1} \exp\left\{-\beta\left(d - \sum_{i=1}^{n} \log(x_i)\right)\right\} \mathbf{1}_{(0,+\infty)}(\beta).$$

   Se $d > \sum_{i=1}^n \log(x_i)$, il membro di destra in quest'ultima diseguaglianza è, evidentemente, il kenel di una densità gamma di parametri shape pari a $n + c$ e rate $d - \sum_{i=1}^n \log(x_i)$.

∎

STATISTICA BAYESIANA A. Guglielmi, R. Argiento 24.09.2014

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 Si consideri il seguente modello

$$X_1, \dots, X_n | \mu, \tau \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \frac{1}{\tau}) \tag{1}$$

$$(\mu, \tau) \sim \pi(\mu|\tau) \times \pi(\tau) = \mathcal{N}(\mu_0, \frac{1}{n_0 \tau}) \times gamma(\alpha, \beta) \tag{2}$$

dove $(\mu, \tau) \in \Theta = \mathbb{R} \times (0, +\infty)$, e n_0, α e β sono iperparametri positivi, μ_0 iperparametro reale. In sintesi, il modello è gaussiano con media μ e varianza $1/\tau$ con iniziale per (μ, τ) di tipo normale-gamma.

1. Si ricavi la distribuzione finale di (μ, τ), sulla base di un campione osservato $\underline{x} = (x_1, \dots, x_n)$. *Suggerimento*: potrà essere utile adottare la seguente notazione

$$\mu_n := \frac{\sum_i x_i + n_0 \mu_0}{n + n_0}, \quad K_n := \frac{\sum_i x_i^2 + n_0 \mu_0^2}{n + n_0} - \mu_n^2.$$

2. Verificare che $K_n \geq 0$.

3. La distribuzione $\pi(\mu, \tau)$ in (2) è coniugata al modello (1)? Giustificare le conclusioni raggiunte.

4. Ricavare le distribuzioni finale e iniziale del solo parametro μ (i.e. la marginale di μ, a priori e a posteriori), descrivendo esplicitamente i passaggi nei conti analitici.

5. Proporre un algoritmo Gibbs sampler per fare inferenza a posteriori per (μ, τ), descrivendone esplicitamente i passi.

Sono noti dei dati $x_1, \dots, x_n$, dove x_i è il logaritmo naturale del tempo impiegato (in minuti) per percorrere un giro completo di un circuito in un parco cittadino da una stessa persona in $n = 40$ giornate distinte. In particolare, $\sum x_i = 163.22$ e $\sum_i x_i^2 = 1670.55$.

6. Si calcolino valore atteso e varianza a posteriori per la media delle durate μ, sulla base dei dati a disposizione e del modello (1)-(2); si scelgano gli iperparametri $\alpha > 1$, β e μ_0 in modo che i seguenti vincoli sui momenti della marginale unidimensionale dei dati siano soddisfatti:

$$\mathrm{E}[X_1] = 3.95, \quad \mathrm{Var}[X_1] = 0.1,$$

con $n_0 = 10$. Si discuta la robustezza delle quantità a posteriori ricavate. *Suggerimento: vi sarà utile utilizzare le proprietà della speranza matematica condizionale.*

Ricordate inoltre che se $\tau \sim gamma(\alpha, \beta)$ e $\alpha > 1$, allora

$$\mathrm{E}\left(\frac{1}{\tau}\right) = \frac{\beta}{\alpha - 1}.$$

Soluzione Da (1) si ricava la verosimiglianza (con $\theta = (\mu, \tau)$):

$$L(\theta; x_1, \dots, x_n) = \left(\prod_{i=1}^{n} \sqrt{\frac{\tau}{2\pi}} \mathrm{e}^{-\frac{\tau (x_i - \mu)^2}{2}} \right) \propto \tau^{n/2} \mathrm{e}^{-\frac{\tau}{2} \sum_1^n (x_i - \mu)^2}, \ \tau > 0, \mu \in \mathbb{R}.$$

1. La distribuzione finale di μ, τ si ricava dal Teorema di Bayes

$$\begin{aligned}\pi(\mu, \tau|\underline{x}) &\propto L(\theta; \underline{x})\pi(\mu|\tau) \times \pi(\tau)\\ &\propto \tau^{n/2}\mathrm{e}^{-\frac{\tau}{2}\sum_1^n (x_i-\mu)^2}\sqrt{\frac{n_0\tau}{2\pi}}\mathrm{e}^{-\frac{n_0\tau}{2}(\mu-\mu_0)^2}\tau^{\alpha-1}\mathrm{e}^{-\beta\tau}\\ &\propto \tau^{n/2+\alpha-1}\mathrm{e}^{-\beta\tau}\tau^{1/2}\exp\left(-\frac{1}{2}(\tau\sum_1^n(x_i-\mu)^2 + n_0\tau(\mu-\mu_0)^2)\right), \ \tau > 0, \mu \in \mathbb{R}.\end{aligned}$$

Da

$$\begin{aligned}\tau\sum_1^n(x_i-\mu)^2 + n_0\tau(\mu-\mu_0)^2 &= (n+n_0)\tau\left(\mu^2 - 2\mu\frac{\sum x_i + n_0\mu_0}{n+n_0} + \frac{\sum x_i^2 + n_0\mu_0^2}{n+n_0}\right)\\ &= (n+n_0)\tau\left((\mu - \frac{\sum x_i + n_0\mu_0}{n+n_0})^2 + \frac{\sum x_i^2 + n_0\mu_0^2}{n+n_0} - (\frac{\sum x_i + n_0\mu_0}{n+n_0})^2\right)\\ &= (n+n_0)\tau\left((\mu-\mu_n)^2 + K_n\right)\end{aligned}$$

si trova

$$\pi(\mu, \tau|\underline{x}) \propto \tau^{n/2+\alpha-1}\exp\left[-\tau\left(\frac{(n+n_0)K_n}{2} + \beta\right)\right]\tau^{1/2}\exp\left(-\frac{(n+n_0)\tau}{2}(\mu-\mu_n)^2\right), \ \tau > 0, \mu \in \mathbb{R};$$

dunque

$$\begin{aligned}\pi(\mu, \tau|\underline{x}) &= \pi(\mu, |\tau, \underline{x}) \times \pi(\tau, |\underline{x})\\ &= \mathcal{N}\left(\mu_n, \frac{1}{(n+n_0)\tau}\right) \times gamma\left(\frac{n}{2} + \alpha, \frac{(n+n_0)K_n}{2} + \beta\right).\end{aligned}$$

2. K_n rappresenta la varianza di una distribuzione discreta uniforme con supporto $\{x_1, \ldots, x_n, \mu_0\}$ con pesi rispettivi $\{1/(n+n_0), \ldots, 1/(n+n_0), n_0/(n+n_0)\}$ e dunque è quantità non-negativa.

3. Dal punto 1. si vede che la prior normale-gamma (2) è coniugata al modello (1) perché la distribuzione finale è ancora di tipo normale-gamma.

4. Si tratta si ricavare

$$\pi(\mu|\underline{x}) = \int_0^{+\infty}\pi(\mu, \tau|\underline{x})d\tau \quad \text{e} \quad \pi(\mu) = \int_0^{+\infty}\pi(\mu, \tau)d\tau.$$

Per la proprietà di coniugio, le due distribuzioni si ottengono con conti analoghi. Indichiamo genericamente con il pedice 1, $\bullet_1$, il parametro generico: poi adatteremo al caso della marginale a priori e della marginale a posteriori. Dobbiamo calcolare l'integrale, rispetto a τ che varia in $(0, +\infty)$, della seguente funzione (di τ)

$$\mathcal{N}\left(\cdot; \mu_1, \frac{1}{n_1\tau}\right) \times gamma\,(\tau; \alpha_1, \beta_1)\,;$$

dunque

$$\int_0^{+\infty}\sqrt{\frac{n_1\tau}{2\pi}}\mathrm{e}^{-\frac{n_1\tau}{2}(\mu-\mu_1)^2}\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\tau^{\alpha_1-1}\mathrm{e}^{-\beta_1\tau}d\tau = \sqrt{\frac{n_1}{2\pi}}\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\int_0^{+\infty}\tau^{\alpha_1+1/2-1}\mathrm{e}^{-\tau(\beta_1+\frac{n_1}{2}(\mu-\mu_1)^2)}d\tau$$

e, riconoscendo il nucleo di una densità gamma,

$$= \sqrt{\frac{n_1}{2\pi}}\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)}\frac{\Gamma(\alpha_1)}{(\beta_1 + \frac{n_1}{2}(\mu-\mu_1)^2)^{\alpha_1+1/2}} = \sqrt{\frac{n_1}{2\pi\beta_1}}\frac{\Gamma(\alpha_1+1/2)}{\Gamma(\alpha_1)}\left(1 + \frac{n_1\alpha_1/\beta_1}{2\alpha_1}(\mu-\mu_1)^2\right)^{-\frac{2\alpha_1+1}{2}},$$

che è una distribuzione

$$t-\textit{univariata}\left(\mu_1, \frac{\beta_1}{n_1\alpha_1}, 2\alpha_1\right) \text{ con } 2\alpha_1 \text{ gradi di libertà, "media" } \mu_1 \text{ e "varianza" } \frac{\beta_1}{n_1\alpha_1}.$$

Ricordiamo che se $2\alpha_1 > 1$, μ_1 è effettivamente la media di tale distribuzione, mentre, se $2\alpha_1 > 2$, $\beta_1/(n_1(\alpha_1 - 1))$ è la varianza.
Per la distribuzione finale, $n_1 = n + n_0$, $\mu_1 = \mu_n$, $\alpha_1 = \alpha + n/2$ e $\beta_1 = \beta + (n + n_0)K_n/2$, e quindi $\pi(\mu|\underline{x})$ è

$$t-\text{univariata}\left(\mu_n, \frac{2\beta + (n+n_0)K_n}{(n+n_0)(2\alpha+n)}, 2\alpha + n\right),$$

mentre per la distribuzione iniziale $n_1 = n_0$, $\mu_1 = \mu_0$, $\alpha_1 = \alpha$ e $\beta_1 = \beta$, e quindi $\pi(\mu)$ è

$$t-\text{univariata}\left(\mu_0, \frac{\beta}{n_0\alpha}, 2\alpha\right).$$

5. Nel Gibbs sampler si deve campionare iterativamente dalle distribuzioni *full conditionals*:

$$-\ \mathcal{L}(\mu|\tau, \underline{x}) = \mathcal{N}\left(\mu_n, \frac{1}{(n+n_0)\tau}\right)$$

$$-\ \mathcal{L}(\tau|\mu, \underline{x}) \propto L(\mu, \tau; \underline{x})\pi(\mu|\tau)\pi(\tau) \propto \tau^{(n+1)/2+\alpha-1} e^{-\beta\tau} e^{-\frac{n_0\tau}{2}(\mu-\mu_0)^2} e^{-\frac{\tau}{2}\sum_i (x_i-\mu)^2}, \ \tau > 0,$$

$$= \textit{gamma}\left(\alpha + \frac{n+1}{2}, \beta + \frac{n_0}{2}(\mu-\mu_0)^2 + \frac{1}{2}\sum_i (x_i - \mu)^2\right).$$

6. Applicando le note proprietà della speranza condizionale si trova:

$$\mathrm{E}(X_1) = \mathrm{E}[\mathrm{E}(X_1|\mu,\tau)] = \mathrm{E}[\mu] = \mu_0 = 3.95$$

$$\mathrm{Var}(X_1) = \mathrm{E}[\mathrm{Var}(X_1|\mu,\tau)] + \mathrm{Var}[\mathrm{E}(X_1|\mu,\tau)] = \mathrm{E}\left[\frac{1}{\tau}\right] + \mathrm{Var}[\mu] = \mathrm{E}\left[\frac{1}{\tau}\right] + \frac{1}{n_0}\mathrm{E}\left[\frac{1}{\tau}\right]$$

$$= \left(1 + \frac{1}{n_0}\right)\mathrm{E}\left[\frac{1}{\tau}\right] = \frac{n_0+1}{n_0}\frac{\beta}{\alpha-1} = 0.1, \quad \alpha > 1.$$

Pertanto dall'ultima equazione si ricava $\beta/(\alpha - 1) = 1/11$, e in conclusione

$$\mu_0 = 3.95, n_0 = 10, \beta = \frac{\alpha-1}{11}, \alpha > 1.$$

Dunque

$$\mathrm{E}_\pi[\mu|\underline{x}] = \mu_n = 4.0544, \quad K_n = 36.5315 - (\mu_n)^2 = 20.09334,$$

$$\mathrm{Var}_\pi[\mu|\underline{x}] = \frac{\beta + \frac{(n+n_0)K_n}{2}}{(n+n_0)(\alpha + n/2 - 1)} = \frac{\beta + 25K_n}{50(\alpha+19)} = \frac{\alpha + 5524.669}{550\alpha + 10450}.$$

La media a posteriori non dipende da (α, β), mentre la varianza decresce per α crescente ($\alpha > 1$) e varia in $(1/550, 0.50233]$.

∎

STATISTICA BAYESIANA A. Guglielmi, R. Argiento 12.09.2014

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 Sia $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$ una collezione di punti di $\mathbb{R}^2$. Si dice che $\mathcal{X}$ è un processo di Bernoulli di parametro $\Lambda(\cdot)$, e si scrive $\mathcal{X}$ ~PB($\Lambda(\cdot)$), se

$$\begin{array}{rcl} \boldsymbol{X}_1, \ldots, \boldsymbol{X}_N | N & \stackrel{\text{iid}}{\sim} & \dfrac{\Lambda(\cdot)}{\Lambda^*} \\ N | \Lambda^* & \sim & Poisson(\Lambda^*) \end{array} \quad (1)$$

dove $\Lambda(\cdot)$ è una misura finita su $\mathbb{R}^2$ e $\Lambda^* := \Lambda(\mathbb{R}^2)$.

Si vuole monitorare il tasso del numero di imbarcazioni in un certo tratto di mare, cioè il numero di imbarcazioni per superficie infinitesima. Naturalmente si suppone che un tratto di mare si possa rappresentare con una superficie piana. Al fine di stimare il tasso del numero di imbarcazioni, la capitaneria di porto utilizza dei radar che monitorano delle superfici circolari. Per modellizzare la posizione e il numero di imbarcazioni censite col radar, si utilizza il seguente modello: il numero e la posizione delle barche presenti su una superficie circolare di raggio $\sqrt{\theta}$ (in Km) segue un processo di Bernoulli di parametro $\Lambda(\cdot, a, \theta)$, come in (1), con $\Lambda(\cdot; a, \theta)$ misura costante sulla circonferenza di raggio $\sqrt{\theta}$ su R^2, cioè

$$\Lambda(A; a, \theta) = \int_A a \mathbf{1}_{\mathcal{S}}(\boldsymbol{x}) d\boldsymbol{x}, \quad (2)$$

dove $\mathcal{S} = \{x \in \mathbb{R}^2 : \sqrt{x_1^2 + x_2^2} \leq \sqrt{\theta}\} = \{x \in \mathbb{R}^2 : ||x||^2 \leq \theta\}$ e $a > 0$.

1. Calcolare $\Lambda^* = \Lambda(\mathbb{R}^2)$ e scrivere la verosimiglianza del modello (1)-(2), condizionatamente ad a e θ.
2. Si vuole fare inferenza solo sul parametro a (tasso del numero di imbarcazioni), supponendo θ fissato e noto. Si scriva la distribuzione coniugata ad a, e si calcoli l'aggiornamento dei parametri a posteriori, sulla base di (1).
3. In un esperimento precedente, condotto in un tratto di mare con le stesse caratteristiche di quello in esame, in una circonferenza di area 28.27 km^2 sono state contate 30 barche. Utilizzare il principio del campione equivalente per fissare gli iperparametri della prior coniugata, che verranno adottati per tutto il resto dell'esercizio.
4. Viene monitorato mediante il radar il numero di barche presenti in in tratto di mare di raggio $\sqrt{\theta} = 2$ Km e si contano $n = 11$ barche. Fornire una stima puntuale ed intervallare (di livello 0.95) per a a posteriori, sulla base dei dati (forniti in questo punto) e del modello (1)-(2).
5. Sia N^* il numero di barche che saranno contate nella prossima scannerizzazione in un tratto di mare, con le stesse caratteristiche del precedente, di raggio $\sqrt{\theta}$ Km. Rispondere alle seguenti domande, fornendo risultati in funzione dell'area $|\mathcal{S}|$=$\pi\theta$.
   (a) Quante navi vi aspettate di contare?
   (b) Con quale varianza?
   (c) Scrivete esplicitamente la distribuzione predittiva di N^*. Riconoscete una distribuzione nota? Con quali parametri?

(Per rispondere ai punti (5.a) e (5.b), vi sarà utile utilizzare le proprietà della speranza matematica condizionale.)

Soluzione

1. Osserviamo che l'area della sfera di raggio $\sqrt{\theta}$ è $|\mathcal{S}| = \pi\theta$. Pertanto
$$\Lambda^* = \Lambda(\mathbb{R}^2) = a \int_{\mathbb{R}^2} \mathbf{1}_{\mathcal{S}}(\boldsymbol{x}) d\boldsymbol{x} = a\pi\theta.$$
La densità della misura di probabilità $\Lambda(\cdot;\theta)/\Lambda^*$ è
$$\lambda(\boldsymbol{x}) = \frac{a \mathbf{1}_{\mathcal{S}}(\boldsymbol{x})}{\Lambda^*} = \frac{1}{\pi\theta}\mathbf{1}_{\mathcal{S}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^2.$$
Se $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ è una realizzazione da un PB di parametro $\Lambda(\cdot)$ come in (1)-(2), allora:
$$\begin{aligned} f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, n; a, \theta) &= f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | N = n; a, \theta) \times f(n; a, \theta) \\ &= \prod_{j=1}^{n} \left\{ \frac{1}{\pi\theta} \mathbf{1}_{\mathcal{S}}(\boldsymbol{x}_i) \right\} \times \frac{\mathrm{e}^{-a\pi\theta}(a\pi\theta)^n}{n!} = \frac{1}{n!} a^n \mathrm{e}^{-\pi\theta a} \mathbf{1}_{(||\boldsymbol{x}_{(n)}||, \infty)}(\theta) , \end{aligned} \tag{3}$$
dove $a > 0$ e con $||\boldsymbol{x}_{(n)}||$ abbiamo indicato il massimo fra $||\boldsymbol{x}_1||, \ldots, ||\boldsymbol{x}_n||$; per convenzione poniamo $||\boldsymbol{x}_{(0)}|| = 0$.

2. È evidente dalla (3) che la distribuzione gamma è coniugata al parametro a, quando θ è noto. Sia $a \sim gamma(\alpha, \beta)$ allora
$$\pi(a; x_1, \ldots, x_n; \theta) \propto a^{n+\alpha-1} \mathrm{e}^{-(\pi\theta+\beta)a} \mathbf{1}_{(0,+\infty)}(a).$$
Quindi a posteriori $a \sim gamma(\alpha_1, \beta_1)$, con $\alpha_1 = \alpha + n$ e $\beta_1 = |\mathcal{S}| + \beta = \pi\theta + \beta$.

3. Poiché
$$\mathrm{E}(a | x_1, \ldots, x_n; \theta) = \frac{n + \alpha}{|\mathcal{S}| + \beta} = \frac{\beta}{|\mathcal{S}| + \beta} \frac{\alpha}{\beta} + \frac{|\mathcal{S}|}{|\mathcal{S}| + \beta} \frac{n}{|\mathcal{S}|},$$
si ricava che α rappresenta la numerosità di un ipotetico campione equivalente, mentre β rappresenta l'area della circonferenza dove sarebbe stato osservato il campione equivalente. Utilizzando i dati forniti dall'esercizio poniamo $\alpha = 30$ e $\beta = 28.27$.

4. Dai conti fatti al punto precedente, visto che $n = 11$ e $\theta = 4$, si trova $\alpha_1 = 41$, e $\beta_1 = 28.27 + \pi(2)^2 \simeq 28.27 + 12.56637 = 40.83637$. Dunque, una stima bayesiana di a è la media a posteriori, che vale $\hat{a} = \mathrm{E}(a | x_1, \ldots, x_n; \theta) = \frac{\alpha_1}{\beta_1} = 1.00401$.

   Per quanto riguarda la stima intervallare, dato che α_1 è intero, si può scrivere che $(2\beta_1 a) \sim gamma(2\alpha_1/2, 1/2) = \chi^2_{2\alpha_1}$. Quindi $\left(\frac{\chi^2_{2\alpha_1, 0.025}}{2\beta 1}, \frac{\chi^2_{2\alpha_1, 0.975}}{2\beta 1} \right)$ è l'intervallo di credibilità desiderato per θ. Utilizzando le tavole ($\chi^2_{2\alpha_1, 0.025} = 58.84462$ and $\chi^2_{2\alpha_1, 0.975} = 108.9373$) si ottiene l'intervallo $(0.72049, 1.33383)$. Se si utilizza l'approssimazione della $gamma(\alpha_1, \beta_1)$ ad una gaussiana $\mathcal{N}(1.00401, 0.02459)$, si ottiene l'intervallo $(0.69666, 1.31119)$.

5. Sia $X_1^*, \ldots, X_{N^*}^*$ una nuova realizzazione dal processo di Bernoulli; da (1) si deduce che $\mathcal{L}(N^* | a; \theta)$ è una Poisson(Λ^*), dove $\Lambda^* = a|\mathcal{S}|$. Si osservi ora che
$$\begin{aligned} \mathcal{L}(N^* | X_1, \ldots, X_N, N; \theta) &= \int_0^\infty \mathcal{L}(N^* | a; \theta) \pi(a | X_1, \ldots, X_N, N = n; \theta) da \\ &= \int_0^\infty \frac{1}{n^*!} (|\mathcal{S}| a)^{n^*} \mathrm{e}^{-|\mathcal{S}| a} \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} a^{\alpha_1 - 1} \mathrm{e}^{-\beta_1 a} da \\ &= \frac{\beta_1^{\alpha_1}}{n^*! \Gamma(\alpha_1)} |\mathcal{S}|^{n^*} \frac{\Gamma(n^* + \alpha)}{(|\mathcal{S}| + \beta_1)^{n^* + \alpha_1}} \\ &= \frac{\Gamma(n^* + \alpha_1)}{n^*! \Gamma(\alpha_1)} \left(\frac{\beta_1}{|\mathcal{S}| + \beta_1} \right)^{\alpha_1} \left(\frac{|\mathcal{S}|}{|\mathcal{S}| + \beta_1} \right)^{n^*}, \quad n^* = 0, 1, 2, \ldots. \end{aligned}$$

In quest'ultima riga riconosciamo una distribuzione binomiale negativa di parametri $p = \left(\frac{\beta_1}{|\mathcal{S}|+\beta_1}\right)$ e $b = \alpha_1$. E questa è la risposta alla domanda 5(c).

Ricordiamo che la media e la varianza di una binomiale negativa di parametri b e p sono, rispettivamente, $b\frac{1-p}{p}$ e $b\frac{1-p}{p^2}$. Allora banalmente

(a) $\mathrm{E}(N^*|X_1, \ldots, X_N, N) = \frac{\alpha_1}{\beta_1}|\mathcal{S}|$

(b) $\mathrm{Var}(N^*|X_1, \ldots, X_N, N) = \frac{\alpha_1}{\beta_1^2}|\mathcal{S}|(|\mathcal{S}| + \beta_1)$.

Se non ricordiamo la media e la varianza di una binomiale negativa a memoria, allora basta osservare che

(a)

$$\begin{aligned}
\mathrm{E}(N^*|X_1, \ldots, X_N, N; \theta) &= \mathrm{E}(\mathrm{E}(N^*|a, X_1, \ldots, X_N, N)|X_1, \ldots, X_N, N) \\
&= \mathrm{E}(\mathrm{E}(N^*|a)|X_1, \ldots, X_N, N) \\
&= \mathrm{E}(|\mathcal{S}|a|X_1, \ldots, X_N, N) \\
&= \frac{\alpha_1}{\beta_1}|\mathcal{S}|
\end{aligned}$$

(b)

$$\begin{aligned}
\mathrm{Var}(N^*|X_1, \ldots, X_N, N; \theta) &= \mathrm{Var}(\mathrm{E}(N^*|a, X_1, \ldots, X_N, N)|X_1, \ldots, X_N, N) \\
&\quad + \mathrm{E}(\mathrm{Var}(N^*|a, X_1, \ldots, X_N, N)|X_1, \ldots, X_N, N) \\
&= \mathrm{Var}(|\mathcal{S}|a|X_1, \ldots, X_N, N) + \mathrm{E}(|\mathcal{S}|a|X_1, \ldots, X_N, N) \\
&= \frac{\alpha_1}{\beta_1^2}|\mathcal{S}|^2 + \frac{\alpha_1}{\beta_1}|\mathcal{S}| \\
&= \frac{\alpha_1}{\beta_1^2}|\mathcal{S}|(|\mathcal{S}| + \beta_1).
\end{aligned}$$

∎

STATISTICA BAYESIANA A. Guglielmi e R. Argiento 09.07.2014

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 I dati in Tabella 1 mostrano le intenzioni di voto di un campione di 1000 elettori italiani che hanno assistito ad entrambi i comizi conclusivi dei leader delle coalizioni di Sinistra e di Destra, relativi alle ultime elezioni europee.

Favorevoli	Sinistra	Destra	Totale
prima dei comizi	491	509	1000
dopo i comizi	523	477	1000

Tabella 1: Favorevoli alla coalizione di Sinistra o a quella di Destra.

Si assuma che i dati, prima e dopo i comizi, siano campioni condizionatamente indipendenti, da due distribuzioni di Bernoulli di parametri p e q, rispettivamente (p è la probabilità che un elettore intenda votare la coalizione di Sinistra prima dei comizi conclusivi, e q è la probabilità che un elettore intenda votare la coalizione di Sinistra dopo i comizi). Come distribuzione iniziale per (p, q) si assuma:

$$p, q|Z = z \overset{\text{iid}}{\sim} beta(z, 1) \qquad Z \sim gamma(\alpha, \beta), \tag{1}$$

dove $\alpha, \beta > 0$ sono tali che $\mathrm{E}(Z) = \mathrm{Var}(Z) = 1$.

1. Calcolare la distribuzione finale di (p, q). I parametri p e q sono a posteriori indipendenti? (*Suggerimento:* non è necessario sostituire i valori numerici di x_i e y_i, $i = 1, \ldots, n$ in questa risposta.)

2. Descrivendone esplicitamente i passi, proporre un algoritmo MCMC per valutare le probabilità a posteriori $\mathbb{P}(p < q|\text{dati})$ e $\mathbb{P}(q > 0.5|\text{dati})$.

Supponiamo per il resto dell'esercizio che $Z = 1$ q.c. nella prior (1) (invece che Z con distribuzione *gamma*).

3. Calcolare la corrispondente distribuzione a posteriori di (p, q), sulla base dei dati in Tabella 1. Ricavare le corrispondenti media e varianza a posteriori di p e di q.

4. Calcolare le probabilità $\mathbb{P}(p < q|\text{dati})$ e $\mathbb{P}(q > 0.5|\text{dati})$, utilizzando una opportuna approssimazione della distribuzione finale di (p, q) determinata al punto 3.

Si ricordi che se $U \sim beta(\alpha, \beta)$, allora $\mathrm{E}(U) = \frac{\alpha}{\alpha+\beta}$, $\mathrm{Var}(U) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Soluzione

1. Il parametro è (p, q, Z). Calcoliamo anzitutto la sua posterior, e poi calcoleremo la marginale a posteriori del solo (p, q).

   Condizionatamente a $(p, q, Z) \in (0, 1) \times (0, 1) \times (0, +\infty)$, le v.a. X_i e Y_i, $i = 1, \ldots, n$, con $n = 1000$, sono di Bernoulli di parametro p e q rispettivamente, e tutte indipendenti tra loro. Si noti che

$$\mathcal{L}(X_1, \ldots, X_n, Y_1, \ldots, Y_n|p, q, Z) = \mathcal{L}(X_1, \ldots, X_n, Y_1, \ldots, Y_n|p, q)$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \prod_{i=1}^{n} q^{y_i}(1-q)^{1-y_i} = p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i} q^{\sum_1^n y_i}(1-q)^{n-\sum_1^n y_i},$$

con $p, q \in (0,1)$. Invece la densità a priori $\pi(p,q,z)$ è:

$$\pi(p,q,z) = \pi(p,q|Z=z)\pi_Z(z) = zp^{z-1}\mathbf{1}_{(0,1)}(p) zq^{z-1}\mathbf{1}_{(0,1)}(q)\mathrm{e}^{-z}\mathbf{1}_{(0,+\infty)}(z)$$
$$= z^2\mathrm{e}^{-z}p^{z-1}q^{z-1}\ \mathbf{1}_{(0,1)}(p)\mathbf{1}_{(0,1)}(q)\mathbf{1}_{(0,+\infty)}(z),$$

visto che $1 = \mathrm{E}(Z) = \alpha/\beta$ e $1 = \mathrm{Var}(Z) = \alpha/\beta^2$, cioè $\alpha = \beta = 1$.

Pertanto, la distribuzione finale di p, q, Z è

$$\pi(p,q,z|\boldsymbol{X},\boldsymbol{Y}) = \frac{\mathcal{L}(X_1,\ldots,X_n,Y_1,\ldots,Y_n|p,q,Z)\pi(p,q|Z=z)\pi_Z(z)}{\int \mathcal{L}(X_1,\ldots,X_n,Y_1,\ldots,Y_n|p,q,Z)\pi(p,q|Z=z)\pi_Z(z)dpdqdz}$$
$$= \frac{p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}q^{\sum_1^n y_i}(1-q)^{n-\sum_1^n y_i}z^2\mathrm{e}^{-z}p^{z-1}q^{z-1}}{\int_0^{+\infty} dz\ \mathrm{e}^{-z}z^2\int_0^1\int_0^1 p^{z+\sum_1^n x_i-1}(1-p)^{n-\sum_1^n x_i+1-1}q^{z+\sum_1^n y_i-1}(1-q)^{n-\sum_1^n y_i+1-1}dpdq}$$
$$= \frac{z^2\mathrm{e}^{-z}p^zq^z\ p^{\sum_1^n x_i-1}(1-p)^{n-\sum_1^n x_i}q^{\sum_1^n y_i-1}(1-q)^{n-\sum_1^n y_i}}{m(\boldsymbol{x},\boldsymbol{y})}\mathbf{1}_{(0,1)}(p)\mathbf{1}_{(0,1)}(q)\mathbf{1}_{(0,+\infty)}(z),$$

dove

$$m(\boldsymbol{x},\boldsymbol{y}) = \Gamma(n-\sum x_i+1)\Gamma(n-\sum y_i+1)\int_0^{+\infty} z^2\mathrm{e}^{-z}\frac{\Gamma(z+\sum x_i)}{\Gamma(z+n+1)}\frac{\Gamma(z+\sum y_i)}{\Gamma(z+n+1)}dz.$$

Quindi, la distribuzione finale di p, q si ottiene integrando rispetto a z la densità $\pi(p,q,z|\boldsymbol{X},\boldsymbol{Y})$:

$$\pi(p,q|\boldsymbol{X},\boldsymbol{Y}) = \frac{p^{\sum_1^n x_i-1}(1-p)^{n-\sum_1^n x_i}q^{\sum_1^n y_i-1}(1-q)^{n-\sum_1^n y_i}\int_0^{+\infty} z^2\mathrm{e}^{-z}\mathrm{e}^{z\log(pq)}dz}{m(\boldsymbol{x},\boldsymbol{y})}$$
$$= \frac{p^{\sum_1^n x_i-1}(1-p)^{n-\sum_1^n x_i}q^{\sum_1^n y_i-1}(1-q)^{n-\sum_1^n y_i}\int_0^{+\infty} z^{3-1}\mathrm{e}^{-z(1-log(pq))}dz}{m(\boldsymbol{x},\boldsymbol{y})}$$
$$= \frac{2}{m(\boldsymbol{x},\boldsymbol{y})}\frac{1}{(1-\log(pq))^3}p^{\sum_1^n x_i-1}(1-p)^{n-\sum_1^n x_i}q^{\sum_1^n y_i-1}(1-q)^{n-\sum_1^n y_i}\mathbf{1}_{(0,1)}(p)\mathbf{1}_{(0,1)}(q)$$

Ovviamente p e q NON sono a posteriori indipendenti.

In alternativa, si può caratterizzare $\pi(p,q|\boldsymbol{X},\boldsymbol{Y})$ nel seguente modo:

$$\pi(p,q|\boldsymbol{X},\boldsymbol{Y}) = \int_0^{+\infty}\pi(p,q|Z,\boldsymbol{X},\boldsymbol{Y})\pi(Z|\boldsymbol{X},\boldsymbol{Y})dZ.$$

Il primo fattore nel prodotto a destra corrisponde alla posterior del modello equivalente che si ottiene fissando Z (condizionatamente a Z), ed è facile verificare che:

$$\pi(p,q|Z,\boldsymbol{X},\boldsymbol{Y}) \propto p^{z+\sum_1^n x_i-1}(1-p)^{n-\sum_1^n x_i+1-1}q^{z+\sum_1^n y_i-1}(1-q)^{n-\sum_1^n y_i+1-1}$$
$$= beta(z+\sum_1^n x_i, n-\sum_1^n x_i+1)\times beta(z+\sum_1^n y_i, n-\sum y_i+1).$$

Il secondo fattore, invece, è:

$$\pi(Z|\boldsymbol{X},\boldsymbol{Y}) \propto \mathcal{L}(Z,\boldsymbol{X},\boldsymbol{Y}) = \int dp\int dq\ \mathcal{L}(Z,p,q,\boldsymbol{X},\boldsymbol{Y})$$
$$= \int \mathcal{L}(\boldsymbol{X},\boldsymbol{Y}|Z,p,q)\pi(p,q|z)\pi_Z(z)dpdq$$
$$= \int z^2\mathrm{e}^{-z}p^{z+\sum_1^n x_i-1}(1-p)^{n-\sum_1^n x_i+1-1}q^{z+\sum_1^n y_i-1}(1-q)^{n-\sum_1^n y_i+1-1}dpdq$$
$$= \Gamma(n-\sum x_i+1)\Gamma(n-\sum y_i+1)\ z^2\mathrm{e}^{-z}\frac{\Gamma(z+\sum x_i)}{\Gamma(z+n+1)}\frac{\Gamma(z+\sum y_i)}{\Gamma(z+n+1)},\quad z>0.$$

2. Anzitutto, costruiamo, per esempio, un algoritmo Gibbs sampler, per generare dalla posterior $\pi((p,q),z|\boldsymbol{X},\boldsymbol{Y})$, calcolando iterativamente le distribuzioni *full conditionals*:

   - $\mathcal{L}((p,q)|Z=z,\boldsymbol{X},\boldsymbol{Y})$: è la legge che abbiamo già descritto in precedenza, e cioè

   $$\mathcal{L}(p,q|Z=z,\boldsymbol{X},\boldsymbol{Y}) = beta(z+\sum_1^n x_i, n-\sum_1^n x_i+1)\times beta(z+\sum_1^n y_i, n-\sum_1^n y_i+1).$$

   - $\mathcal{L}(Z|(p,q),\boldsymbol{X},\boldsymbol{Y}) \propto \mathcal{L}(\boldsymbol{X},\boldsymbol{Y},p,q,Z) \propto \mathcal{L}(\boldsymbol{X},\boldsymbol{Y}|p,q)\mathcal{L}(p,q|Z)\mathcal{L}(Z) \propto \mathcal{L}(p,q|Z)\mathcal{L}(Z)$.
   Quindi

   $$\begin{aligned}\mathcal{L}(Z|(p,q),\boldsymbol{X},\boldsymbol{Y}) &\propto zp^{z-1}zq^{z-1}\mathrm{e}^{-z},\ z>0\\ &\propto z^2(pq)^z\mathrm{e}^{-z} = z^{3-1}\mathrm{e}^{-z(1-\log(pq))},\ z>0,\end{aligned}$$

   che risulta essere una distribuzione $gamma(3, 1-\log(pq))$.

   Per calcolare, col Gibbs sampler sopra delineato, $\mathbb{P}(p-q<0|\text{dati})$ si genera un numero elevato G di osservazioni di $(p^{(g)},q^{(g)},Z^{(g)})$, simulando, ad ogni iterazione g, in questo modo: genero $Z^{(g)}$ da una $gamma(3,1-\log(p^{(g-1)}q^{(g-1)}))$, e successivamente $(p^{(g)},q^{(g)})$ dalla distribuzione prodotto $beta(Z^{(g)}+491,510)\times beta(Z^{(g)}+523,478)$. La frequenza relativa con cui si presenta l'evento $p^{(g)}<q^{(g)}$ è una stima MCMC della probabilità in esame.

   Riportiamo di seguito uno script R che traduce questo algoritmo:

```
set.seed(1)
G = 10000
theta=matrix(nrow=G,ncol=3) #Inizializzo una matrice con 3 colonne e G righe

# La prima colonna di theta:   p
# la seconda colonna:          q
# la terza colonna:            Z

### Dati
n=1000
SX= 491
SY= 523
# Inizializzo
theta[1,]=c(0.5,0.5,1)

for(s in 2:G) {
theta[s,3] = rgamma(1,shape=3,rate=1-log(theta[s-1,1]*theta[s-1],2))
theta[s,1] = rbeta(1,theta[s,3]+SX, n-SX+1)
theta[s,2] = rbeta(1,theta[s,3]+SY, n-SY+1)
}
stima_MCMC=mean((theta[,1]-theta[,2])<0)
print(stima_MCMC)
[1] 0.9235
```

   Analogamente si trova che una stima MCMC di $\mathbb{P}(q>0.5|\text{dati})$ è 0.9255.

3. Se $Z = 1$ q.c., vuol dire che la prior di (p, q) che ora consideriamo è il prodotto di due $beta(1, 1)$. Ripetendo i conti al punto 1., è facile verificare che

$$p, q|\boldsymbol{X}, \boldsymbol{Y} \sim beta(492, 510) \times beta(524, 478).$$

Quindi ora a posteriori p e q sono indipendenti.

Ricordando le formule di media e varianza di una distribuzione $beta(\alpha, \beta)$, si trova:

$$\begin{aligned} &\mathrm{E}(p|\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}(p|\boldsymbol{X}) = 0.491018 \qquad &&\mathrm{Var}(p|\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{Var}(p|\boldsymbol{X}) = 0.000249 \\ &\mathrm{E}(q|\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}(p|\boldsymbol{Y}) = 0.522954 \qquad &&\mathrm{Var}(p|\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{Var}(p|\boldsymbol{Y}) = 0.000249. \end{aligned}$$

4. Osserviamo che, se $U \sim beta(\alpha, \beta)$ e α, β sono *grandi*, U approssimativamente ha distribuzione $\mathcal{N}(\mu, \sigma^2)$, con $\mu = \mathrm{E}(U) = \frac{\alpha}{\alpha+\beta}$ e varianza $\sigma^2 = \mathrm{Var}(U) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Quindi possiamo ritenere

$$\begin{aligned} p|\text{dati} &\overset{approx}{\sim} \mathcal{N}(\mu_p = 0.491018, \sigma_p^2 = 0.000249) \\ q|\text{dati} &\overset{approx}{\sim} \mathcal{N}(\mu_q = 0.522954, \sigma_q^2 = 0.000249). \end{aligned}$$

Per l'indipendenza a posteriori

$$p - q \overset{approx}{\sim} \mathcal{N}\left(\mu_p - \mu_q = -0.031936, \sigma_p^2 + \sigma_q^2 = 0.000498\right).$$

In conclusione:

$$\mathbb{P}\left(p - q < 0|\text{dati}\right) \simeq \Phi\left(\frac{-(\mu_p - \mu_q)}{\sqrt{\sigma_p^2 + \sigma_q^2}}\right) = \Phi(1.431086) = 0.923797.$$

In modo analogo

$$\mathbb{P}\left(q > 0.5|\text{dati}\right) \simeq 1 - \Phi\left(\frac{0.5 - \mu_q}{\sqrt{\sigma_q^2}}\right) = 1 - \Phi(-1.454651) = \Phi(1.454651) = 0.927117.$$

∎

STATISTICA BAYESIANA A. Guglielmi e R. Argiento 27.02.2014

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 È noto che se X_1 e X_2 sono due variabili aleatorie discrete indipendenti, entrambe con supporto in $\{1, 2, 3 \dots\}$, allora, per ogni k nel supporto,

$$\mathbb{P}(X_1 < X_2, \min\{X_1, X_2\} = k) = \mathbb{P}(X_2 > k)\mathbb{P}(X_1 = k), \tag{1}$$
$$\mathbb{P}(X_2 \leq X_1, \min\{X_1, X_2\} = k) = \mathbb{P}(X_1 \geq k)\mathbb{P}(X_2 = k). \tag{2}$$

1. Si dimostri la formula (1).

Si considerino ora due macchinari analoghi, prodotti da due aziende diverse, con probabilità di guasto pari a $p_i \in (0, 1)$ e sia X_i la variabile aleatoria che indica il tempo (discreto) di guasto del macchinario i, con $i = 1, 2$. Si assume che, condizionatamente a p_i, X_i abbia distribuzione geometrica di parametro p_i, ossia

$$\mathbb{P}(X_i = k|p_i) = (1 - p_i)^{k-1} p_i \qquad k = 1, 2, 3, \dots.$$

2. Verificare che $\mathbb{P}(X_i > k|p_i) = (1 - p_i)^k$, $k = 1, 2, 3, \dots$.

Per determinare quale dei due macchinari sia più affidabile si conduce il seguente esperimento: si considerano n coppie di macchinari dalle due aziende, e per ciascuna coppia si utilizzano entrambi i macchinari fin quando uno dei due non si guasta. Sia, allora,

$$C := \begin{cases} 1 & \text{se } X_1 < X_2 \\ 0 & \text{se } X_1 \geq X_2 \end{cases}$$

la variabile aleatoria che indica se il primo macchinario si è guastato prima del secondo e sia

$$K := \min\{X_1, X_2\}$$

il tempo della prima rottura. Come risultato dell'esperimento si ottengono i dati reali

$$(k_1, c_1), \dots, (k_n, c_n),$$

realizzazione del campione $(K_1, C_1), \dots, (K_n, C_n)$, che si assume iid, condizionatamente a (p_1, p_2). In particolare, l'esperimento confronta $n = 20$ coppie di macchinari: in $n_1 = 7$ casi il macchinario di tipo 1 si è guastato prima del macchinario di tipo 2. Inoltre $\sum_1^n k_i = 60$.

3. Si utilizzino le formule (1) e (2) per scrivere la densità congiunta del vettore $(K_1, C_1), \dots, (K_n, C_n)$, condizionatamente a (p_1, p_2). In pratica, scrivere la verosimiglianza dei dati. Per uniformare la notazione, si indichi con n_0 il numero di volte in cui si è osservato $c_i = 0$ nel campione di ampiezza n.

4. Determinare la prior $\pi(p_1, p_2) = \pi_1(p_1)\pi_2(p_2)$ coniugata alla verosimiglianza ricavata al punto 3. Come si aggiornano i parametri di questa prior? I parametri p_1 e p_2 sono ancora indipendenti a posteriori?

5. Le informazioni a nostra disposizione possono essere riassute nel modo seguente: in un ipotetico esperimento sono stati confrontati $m = 5$ macchinari. In $m_1 = 2$ casi si è rotto prima il macchinario 1, mentre i tempi di primo guasto ipoteticamente registrati sono: $(l_1, \dots, l_5) = (11, 16, 4, 2, 4)$. Utilizzare il principio del campione equivalente per fissare gli iperparametri della prior coniugata e calcolarne esplicitamente l'aggiornamento a posteriori.

6. Utilizzando la posterior ricavata ai punti 4. e 5., fornire lo stimatore bayesiano (perdita quadratica) a posteriori di p_1 e di p_2.

7. Con la medesima posterior, determinare $\mathbb{P}(p_1 < p_2|dati)$, sapendo che se $F_{a,b}$ è la funzione di ripartizione di una distribuzione beta di parametri a e b **interi**, allora

$$F_{a,b}(u) = \sum_{h=0}^{b-1} \binom{a+b-1}{h} u^{a+b-1-h}(1-u)^h.$$

Inoltre, per facilitare i conti numerici vi diciamo che $\log B(16, 81) = -43.807$ e che

$$\log\left\{\sum_{h=0}^{71} \binom{80}{h} B(96-h, 81+h)\right\} = -43.962$$

dove $B(a, b) := \int_0^1 u^{a-1}(1-u)^{b-1}du$.

Soluzione

1. Non è difficile rendersi conto che l'evento $\{X_1 < X_2\} \cap \{\min\{X_1, X_2\} = k\}$ è equivalente all'evento $\{X_1 = k\} \cap \{X_2 > k\}$, dal fatto che X_1 e X_2 sono indipendenti si ricava l'asserto. In modo equivalente:

$$\begin{aligned}
\mathbb{P}(X_1 < X_2, \min\{X_1, X_2\} = k) &= \sum_{x_1=1}^{\infty} \mathbb{P}(X_1 < X_2, \min\{X_1, X_2\} = k|X_1 = x_1)\mathbb{P}(X_1 = x_1) \\
&= \sum_{x_1=1}^{\infty} \mathbb{P}(x_1 < X_2, \min\{x_1, X_2\} = k)\mathbb{P}(X_1 = x_1) \\
&= \sum_{x_1=1}^{\infty} \mathbb{P}(\min\{x_1, X_2\} = k|X_2 > x_1)\mathbb{P}(X_2 > x_1)\mathbb{P}(X_1 = x_1)
\end{aligned}$$

Dato che $\mathbb{P}(\min\{x_1, X_2\} = k|X_2 > x_1) = \mathbf{1}_k(x_1)$, si ottiene l'asserto.

2. Se $X_i \sim geom(p_i)$, allora

$$\begin{aligned}
\mathbb{P}(X_i > k) = \sum_{j=k+1}^{+\infty} p_i(1-p_i)^{j-1} = p_i \sum_{j=0}^{+\infty}(1-p_i)^{j+k} &= p_i(1-p_i)^k \sum_{j=0}^{+\infty}(1-p_i)^j \\
&= p_i(1-p_i)^k \frac{1}{1-(1-p_i)} = (1-p_i)^k.
\end{aligned}$$

3. La verosimiglianza (o distribuzione congiunta condizionale) dei dati può essere scritta come segue

$$\begin{aligned}
L(p_1, p_2; (k_1, c_1), \ldots, (k_n, c_n)) &= \prod_{i:c_i=1} \mathbb{P}(K_i = k_i, C_i = 1) \prod_{i:c_i=0} \mathbb{P}(K_i = k_i, C_i = 0) \\
&= \prod_{i:c_i=1} \mathbb{P}(\min(X_{i1}, X_{i2}) = k_i, X_{i1} < X_{i2}) \prod_{i:c_i=0} \mathbb{P}(\min(X_{i1}, X_{i2}) = k_i, X_{i1} \geq X_{i2}) \\
&= \prod_{i:c_i=1} \{\mathbb{P}(X_{i2} > k_i)\mathbb{P}(X_{i1} = k_i)\} \prod_{i:c_i=0} \{\mathbb{P}(X_{i1} \geq k_i)\mathbb{P}(X_{i2} = k_i)\}.
\end{aligned}$$

Visto che quelle coinvolte sono tutte distribuzioni geometriche, si ha:

$$L(p_1, p_2; (k_1, c_1), \ldots, (k_n, c_n)) = \prod_{i:c_i=1} \left\{(1-p_2)^{k_i} p_1 (1-p_1)^{k_i-1}\right\} \prod_{i:c_i=0} \left\{(1-p_1)^{k_i-1} p_2 (1-p_2)^{k_i-1}\right\}.$$

Se si indica con n_0 in numero di volte in cui si è osservato $c_1 = 0$ e con $n_1 = n - n_0$ il numero restante di prove sulle n totali, allora

$$L(p_1, p_2; (k_1, c_1), \dots, (k_n, c_n)) = \left\{(1-p_2)^{\sum_{i:c_i=1} k_i} p_1^{n_1}(1-p_1)^{\sum_{i:c_i=1} k_i - n_1}\right\}$$
$$\times \left\{(1-p_1)^{\sum_{i:c_i=0} k_i - n_0} p_2^{n0}(1-p_2)^{\sum_{i:c_i=0} k_i - n_0}\right\}.$$

In conclusione si ottiene

$$L(p_1, p_2; (k_1, c_1), \dots, (k_n, c_n)) = p_1^{n_1}(1-p_1)^{\sum_{i=1}^n k_i - n} p_2^{n_0}(1-p_2)^{\sum_{i=1}^n k_i - n_0}.$$

4. Dalla forma della verosimiglianza si evince che, come funzione di p_1 e p_2 la verosimiglianza si fattorizza in due termini, ciascuno dei quali è il kernel di una densità beta. Si ha quindi che, $\pi(p_1)\pi(p_2)$ è il prodotto di due densità beta, $p_1 \sim beta(\alpha_1, \beta_1)$ e $p_1 \sim beta(\alpha_2, \beta_2)$. Applicando il teorema di Bayes si ottiene che

$$\begin{aligned}\pi(p_1, p_2|dati) &\propto L(p_1, p_2; (k_1, c_1), \dots, (k_n, c_n)) \times \pi(p_1, p_2)\\ &\propto p_1^{n_1}(1-p_1)^{\sum_{i=1}^n k_i - n} p_2^{n_0}(1-p_2)^{\sum_{i=1}^n k_i - n_0} \times p_1^{\alpha_1 - 1}(1-p_1)^{\beta_1 - 1}\mathbf{1}_{(0,1)}(p_1) p_2^{\alpha_2 - 1}(1-p_2)^{\beta_2 - 1}\mathbf{1}_{(0,1)}(p_2)\\ &\propto p_1^{n_1 + \alpha_1 - 1}(1-p_1)^{\sum_{i=1}^n k_i - n + \beta_1 - 1}\mathbf{1}_{(0,1)}(p_1) \times p_2^{n_0 + \alpha_2 - 1}(1-p_2)^{\sum_{i=1}^n k_i - n_0 + \beta_2 - 1}\mathbf{1}_{(0,1)}(p_2)\end{aligned}$$

e quindi $p_1 \sim beta(\alpha_{1n}, \beta_{1n})$ dove $\alpha_{1n} = \alpha_1 + n_1, \beta_{1n} = \beta_1 + \sum_{i=1}^n k_i - n$ mentre $p_2 \sim beta(\alpha_{2n}, \beta_{2n})$ dove $\alpha_{2n} = \alpha_2 + n_0, \beta_{2n} = \beta_2 + \sum_{i=1}^n k_i - n_0$. I parametri p_1 e p_2 sono indipendenti a posteriori.

5. Dai conti effettuati al punto precedente, se $(k_1, c_1), \dots, (k_n, c_n)$, è un campione di ampiezza n, allora

$$\begin{aligned}\mathrm{E}(p_1|dati) = \frac{\alpha_1 + n_1}{\beta_1 + \sum_{i=1}^n k_i + \alpha_1 - n_0} &= \frac{\alpha_1}{\alpha_1 + \beta_1} \times \frac{\alpha_1 + \beta_1}{\beta_1 + \sum_{i=1}^n k_i + \alpha_1 - n_0}\\ &+ \frac{n_1}{\sum_{i=1}^n k_i - n_0} \times \frac{\sum_{i=1}^n k_i - n_0}{\beta_1 + \sum_{i=1}^n k_i + \alpha_1 - n_0}.\end{aligned}$$

Di conseguenza, in un ipotetico campione a priori, α_1 rappresenta il numero di volte in cui si è guastato il primo macchinario, e quindi, con la notazione dell'esercizio, $\alpha_1 = m_1 = 2$, mentre di conseguenza $\beta_1 = \sum_{i=1}^m l_i - m_0 - m_1 = \sum_{i=1}^n l_i - m = 37 - 5 = 32$. In modo analogo

$$\begin{aligned}\mathrm{E}(p_2|dati) = \frac{\alpha_2 + n_0}{\beta_2 + \sum_{i=1}^n k_i + \alpha_2} &= \frac{\alpha_2}{\alpha_2 + \beta_2} \times \frac{\alpha_2 + \beta_2}{\beta_2 + \sum_{i=1}^n k_i + \alpha_2}\\ &+ \frac{n_0}{\sum_{i=1}^n k_i} \times \frac{\sum_{i=1}^n k_i}{\beta_2 + \sum_{i=1}^n k_i + \alpha_2}.\end{aligned}$$

Quindi α_2 rappresenta il numero di volte in cui, nel campione ipotetico, non è il macchinario di tipo 1 a rompersi per primo, cioè con la notazione dell'esercizio $\alpha_2 = m_0 = 3$, mentre di conseguenza $\beta_2 = \sum_{i=1}^m l_i - m_0 = 37 - 3 = 34$. A posteriori invece, visto che $n = 20$, $n_1 = 7$, $n_0 = 13$ e $\sum k_i = 60$, si ottiene $\alpha_{1n} = 9$, $\beta_{1n} = 72$, $\alpha_{2n} = 16$ e $\beta_{2n} = 81$.

6. La stima bayesiana a posteriori dei due parametri di interesse è dunque

$$(\widetilde{p}_1, \widetilde{p}_2) = \mathrm{E}((p_1, p_2)|dati) = \left(\frac{\alpha_{1n}}{\alpha_{1n} + \beta_{1n}}, \frac{\alpha_{2n}}{\alpha_{2n} + \beta_{2n}}\right) = \left(\frac{1}{9}, \frac{16}{97}\right).$$

7. Si può scrivere

$$\begin{aligned}\mathbb{P}(p_1 < p_2|dati) &= \int_0^1 \mathbb{P}(p_1 < p_2|p_2 = u, dati) f_{p2|dati}(u)du \\ &= \int_0^1 \mathbb{P}(p_1 < u|dati) f_{p2|dati}(u)du = \int_0^1 F_{\alpha_{1n},\beta_{1n}}(u) f_{p2|dati}(u)du\end{aligned}$$

dove $\mathbb{P}(p_1 < u|dati)$ è la funzione di ripartizione $F_{\alpha_{1n},\beta_{1n}}(u)$ a posteriori di p_1 calcolata in u, mentre

$$f_{p2|dati}(u) = \frac{u^{\alpha_{2n}-1}(1-u)^{\beta_{2n}-1}}{B(\alpha_{2n},\beta_{2n})}\mathbf{1}_{(0,1)}(u)$$

è la densità a posteriori di p_2. Applicando la formula data nel testo nell'esercizio si ottiene

$$F_{\alpha_{1n},\beta_{2n}}(u) = \sum_{h=0}^{\beta_{1n}-1} \binom{\alpha_{1n}+\beta_{1n}-1}{h}(1-u)^h u^{\alpha_{1n}+\beta_{1n}-1-h}.$$

Di conseguenza

$$\begin{aligned}\mathbb{P}(p_1 < p_2|dati) =& \frac{1}{B(\alpha_{2n},\beta_{2n})}\sum_{h=0}^{\beta_{1n}-1}\binom{\alpha_{1n}+\beta_{1n}-1}{h}\int_0^1 u^{\alpha_{2n}+\alpha_{1n}+\beta_{1n}-1-h-1}(1-u)^{\beta_{2n}+h-1}du \\ =& \frac{1}{B(\alpha_{2n},\beta_{2n})}\sum_{h=0}^{\beta_{1n}-1}\binom{\alpha_{1n}+\beta_{1n}-1}{h}B(\alpha_{2n}+\alpha_{1n}+\beta_{1n}-h-1,\beta_{2n}+h).\end{aligned}$$

Sostituendo i valori numerici degli iperparametri a posteriori

$$\begin{aligned}\mathbb{P}(p_1 < p_2|dati) &= \frac{1}{B(16,81)}\sum_{h=0}^{71}\binom{80}{h}B(96-h,81+h) \\ &= \exp\left\{\log\left\{\sum_{h=0}^{71}\binom{80}{h}B(96-h,81+h)\right\} - \log\{B(16,81)\}\right\} \\ &= \exp\{-43.962+43.807\} = \exp\{-0.155\} = 0.856.\end{aligned}$$

∎

STATISTICA BAYESIANA A. Guglielmi e R. Argiento 13.02.2014

Nello svolgere gli esercizi fornire passaggi e spiegazioni: non bastano i risultati finali.

Esercizio 1 Si consideri il seguente modello bayesiano per una successione di variabili aleatorie positive $\mathcal{X} = \{X_n, n = 0, 1, 2, \dots\}$:

$$\begin{cases} X_0 = \omega_0 \\ X_{n+1} = X_n + \omega_{n+1} \quad \text{per } n = 0, 1, 2 \dots \end{cases}$$

dove

$$\omega_0, \omega_1, \dots | \theta \stackrel{iid}{\sim} gamma(1, \theta) \quad \text{e} \quad \theta \sim \pi(\theta).$$

1. Con n generico si determino:
   (a) la distribuzione di X_n condizionatamente a θ;
   (b) la densità condizionale $f_{\boldsymbol{X}|\theta}(x_0, \dots, x_n; \theta)$ di $\boldsymbol{X} := (X_0, X_1, \dots, X_n)$, dato θ;
   (c) l'informazione di Fisher $I_{\boldsymbol{X}}(\theta)$ relativa a tutto il campione $\boldsymbol{X} = (X_0, \dots, X_n)$.

Ricordate la famosa favola di Esopo *La lepre e la tartaruga?* Ora la lepre vuole chiedere la rivincita... Per impostare al meglio la sveglia per la futura gara, la lepre decide di studiare la velocità della tartaruga spiandone gli allenamenti. A tal fine, durante un allenamento della tartaruga, la lepre raccoglie le misurazioni $X_0, X_1, \dots, X_n$ che rappresentano le distanze in metri, dal punto di partenza, percorse dalla tartaruga ogni 10 minuti; quindi X_0 è la distanza percorsa dalla tartaruga nei primi 10 minuti, X_1 quella percorsa nei primi 20 minuti, ecc.. La lepre spia la tartaruga per due ore durante i quali la tartaruga percorre complessivamente 23.62 metri. Utilizzando il modello sopra descritto per tali dati, si risponda alle seguenti domande:

2. Qual è la distribuzione a posteriori di θ se la lepre utilizza la prior di Jeffreys π_1?

3. Qual è la prior π_2 coniugata al modello? Come si aggiornano i suoi parametri a posteriori? A quanto pare, in una precedente gara la tartaruga aveva percorso 18 metri in un'ora. Utilizzare il principio del campione equivalente per fissare i parametri della prior coniugata π_2, sulla base dei dati della gara precedente.

4. Fornire una stima bayesiana (perdita quadratica) a posteriori della funzione di θ che esprime il valor atteso, condizionatamente a θ, della velocità media (in metri al minuto) della tartaruga in un intervallo di tempo di 10 minuti.

La lepre non si fida completamente dell'informazione descritta da π_2 ed elicitata al punto 3. Per rendere più robusta la sua analisi decide di comportarsi in questo modo: tiene fisso il parametro di *rate* β della prior coniugata π_2 (ricavato al punto 3), ma vuole trovare il "miglior" modello al variare del parametro di *shape*, facendo variare quest'ultimo nell'insieme dei naturali positivi.

5. Si indichi con $\kappa \in \{1, 2, \dots\}$ il parametro di *shape* della prior coniugata. Per semplicità di notazione, chiamiamo π_3 questa classe di prior. Si calcoli la densità marginale dei dati $m(\boldsymbol{x}; \kappa, \beta)$ quando π_3 è la prior.

6. Sia $g(\kappa) := m(\boldsymbol{x}; \kappa, \beta)$, quando $\boldsymbol{x}$ è il campione osservato (e β è il parametro di *rate* della prior coniugata ricavato al punto 3). Si trovi il punto di massimo di $g(\kappa)$ per $\kappa \in \{1, 2, \dots\}$.
   Suggerimento: si studi la disequazione $g(\kappa)/g(\kappa + 1) < 1$, la cui soluzione coincide con l'intervallo dei κ per cui g è crescente.

7. Utilizzando come prior π_3, con κ pari al valore ottimo trovato al punto precedente (e β, parametro di *rate*, come al punto 3), si calcoli la probabilità predittiva che in una futura corsa la tartaruga percorra più di 4 metri in mezz'ora.

   Suggerimento: sarà utile ricordare che se k è un intero positivo e $T \sim gamma(k, \theta)$, allora

$$\mathbb{P}(T > x) = \sum_{i=0}^{k-1} \frac{(\theta x)^i}{i!} e^{-\theta x}.$$

Soluzione

1. Vale:

   (a) $X_n = X_{n-1}+\omega_n = X_{n-2}+\omega_n+\omega_{n-1} = \cdots = \omega_0+\omega_1+\cdots+\omega_n$. Condizionatamente a θ, dato che $\omega_0, \omega_1, \ldots, \omega_n \overset{\text{i.i.d.}}{\sim} gamma(1, \theta)$, allora $X_n = \sum_{j=0}^n \omega_j \sim gamma(n+1, \theta)$.

   (b) Le prime $n+1$ osservazioni $X_0, X_1, \ldots, X_n$ non sono indipendenti: utilizzando la regola del prodotto si ha che

$$\mathcal{L}(X_0, \ldots, X_n|\theta) = \mathcal{L}(X_0|\theta)\mathcal{L}(X_1|X_0, \theta) \ldots \mathcal{L}(X_n|X_0, \ldots, X_{n-1}, \theta).$$

   Anzitutto osserviamo che $\mathcal{L}(X_0|\theta)$ è una $gamma(1, \theta)$, mentre per $i = 1, \ldots, n$, $\mathcal{L}(X_i|X_0, \ldots, X_{i-1}, \theta)$ dipende solo da θ e da X_{i-1}. Inoltre

$$\begin{aligned}\mathbb{P}(X_i \leq x_i|X_{i-1} = x_{i-1}, \theta) &= \mathbb{P}(X_{i-1} + \omega_i \leq x_i|\theta) = \mathbb{P}(\omega_i \leq x_i - x_{i-1}|\theta) \\ &= (1 - \mathrm{e}^{-\theta(x_i - x_{i-1})})\mathbf{1}_{(0,\infty)}(x_i - x_{i-1}).\end{aligned}$$

   Quindi

$$f_{X_i|X_{i-1},\theta}(x_i; x_{i-1}, \theta) = \theta \mathrm{e}^{-\theta(x_i - x_{i-1})}\mathbf{1}_{(0,x_{i-1})}(x_i).$$

   Infine, applicando la regola del prodotto alle densità, si ottiene:

$$\begin{aligned}f_{\mathbf{X}|\theta}(x_0, \ldots, x_n; \theta) &= \theta^{n+1}\mathrm{e}^{-\theta x_0}\mathrm{e}^{-\theta(x_1 - x_0)} \cdots \mathrm{e}^{-\theta(x_n - x_{n-1})}\mathbf{1}_{\{0<x_0<x_1<\cdots<x_n\}} \\ &= \theta^{n+1}e^{-\theta x_n}\mathbf{1}_{\{0<x_0<x_1<\cdots<x_n\}}.\end{aligned}$$

   D'ora in poi, se non necessario, sottointenderemo la funzione $\mathbf{1}_{\{0<x_0<x_1<\cdots<x_n\}}$; essa rappresenta la condizione che la successione dei dati $x_0, \ldots, x_1$ sia strettamente crescente, con $x_0 > 0$.

   (c) Poiché $\log f_{\mathbf{X}|\theta}(\boldsymbol{x}; \theta) = (n+1)\log\theta - \theta x_n$, l'informazione di Fisher è il seguente valore atteso:

$$I_{\mathbf{X}}(\theta) = \mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f_{\mathbf{X}|\theta}(\boldsymbol{X}; \theta)\right)^2\right) = \mathrm{E}_\theta\left(\left(\frac{n+1}{\theta} - X_n\right)^2\right) = \mathrm{Var}_\theta(X_n),$$

   visto che che $X_n \sim gamma(n+1, \theta)$, e quindi $\mathrm{E}(X_n) = \frac{n+1}{\theta}$. In conclusione si ha che

$$I_{\mathbf{X}}(\theta) = \frac{n+1}{\theta^2}.$$

   Allo stesso risultato si arriva, in modo più agevole, ricordando che per la famiglia esponenziale si ha:

$$I_{\mathbf{X}}(\theta) = -\,\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f_{\mathbf{X}|\theta}(\boldsymbol{X}; \theta)\right),$$

   e osservando che

$$\frac{\partial^2}{\partial\theta^2}\log f_{\mathbf{X}|\theta}(\boldsymbol{X}; \theta) = -\frac{n+1}{\theta^2}.$$

2. Si ricordi che la prior di Jeffreys π_1 è proporzionale a $\sqrt{I_{\boldsymbol{X}}(\theta)} = \frac{\sqrt{n+1}}{\theta}$ e quindi

$$\pi_1(\theta) \propto \frac{1}{\theta}\mathbf{1}_{(0,+\infty)}(\theta).$$

In questo caso essa è impropria! Ora applicando la regola di Bayes, si trova che

$$\pi_1(\theta|\boldsymbol{x}) \propto f_{\boldsymbol{X}|\theta}(\boldsymbol{x};\theta)\pi_1(\theta) \propto \theta^{n+1-1}\mathrm{e}^{-x_n\theta}\mathbf{1}_{(0,+\infty)}(\theta).$$

Si ha dunque che a posteriori $\pi_1(\cdot|\boldsymbol{x}) = gamma(n+1, x_n)$. Utilizzando i dati raccolti dalla lepre $n = 11$ e $x_n = 23.62$, risulta $\pi_1(\cdot|\boldsymbol{x}) = gamma(n+1, x_n) = gamma(12, 23.62)$.

3. Non è difficile rendersi conto (avendo in mente l'espressione di $f_{\boldsymbol{X}|\theta}$) che la prior coniugata al modello è una gamma. Sia $\pi_2(\theta)$ la densità gamma di parametri α e β. Dal Teorema di Bayes si ottiene facilmente che:

$$\pi_2(\theta|\boldsymbol{x}) \propto \theta^{n+1}e^{-\theta x_n}\theta^{\alpha-1}e^{-\beta\theta}\mathbf{1}_{(0,+\infty)}(\theta) = \theta^{n+1+\alpha-1}e^{-(\beta+x_n)\theta}\mathbf{1}_{(0,+\infty)}(\theta),$$

e pertanto $\pi_2(\theta|\boldsymbol{x}) = gamma(\alpha+n+1, \beta+x_n)$.

Per applicare il principio del campione equivalente $\boldsymbol{y} = (y_0, \ldots, y_{n_0})$, si noti che

$$\mathrm{E}(\theta|\boldsymbol{y}) = \frac{\alpha+n_0+1}{\beta+y_{n_0}} = \frac{\beta}{\beta+y_{n_0}}\frac{\alpha}{\beta} + \frac{y_{n_0}}{\beta+y_{n_0}}\frac{n_0+1}{y_{n_0}}.$$

Da quest'ultima uguaglianza si deduce che α rappresenta n_0+1, mentre β rappresenta y_{n_0}. Utilizzando i dati della gara precedente, sappiamo che $n_0 = 5$ e $y_{n_0} = 18$. Di conseguenza $\alpha = 6$ e $\beta = 18$. Infine, la posterior $\pi_2(\theta|\boldsymbol{x}) = gamma(\alpha_n, \beta_n)$, con $\alpha_n = \alpha+n+1 = 18$ e $\beta_n = \beta + x_n = 41.62$, visto che $n = 11$ e $x_n = 23.62$.

4. Evidentemente, ogni 10 minuti la tartaruga percorre ω metri e, condizionatamente a θ, risulta $\omega \sim gamma(1, \theta)$. La sua velocità media è quindi pari a $\mu := \mathrm{E}_\theta(\omega/10) = 1/(10\theta)$. Uno stimatore bayesiano di μ utilizzando la perdita quadratica è la media a posteriori. Poiché $1/\theta$ si distribuisce come una inversa gamma, $\mathrm{E}(\mu|\boldsymbol{x}) = \frac{\beta_n}{10(\alpha_n-1)} \simeq 0.2448$.

5. Ricordiamo che $f_{\boldsymbol{X}|\theta}(\boldsymbol{x};\theta) = \theta^{n+1}\mathrm{e}^{-\theta x_n}$ (sottointendo l'indicatore); se a priori $\theta \sim gamma(\kappa, \beta)$, allora la densità marginale dei dati è

$$m(\boldsymbol{x};\kappa,\beta) = \frac{\beta^\kappa}{\Gamma(\kappa)}\int_0^{+\infty}\theta^{n+1+\kappa-1}\mathrm{e}^{-\theta(\beta+x_n)}d\theta = \frac{\Gamma(n+1+\kappa)}{\Gamma(\kappa)}\frac{\beta^\kappa}{(\beta+x_n)^{n+1+\kappa}}.$$

6. Utilizzando la notazione $g(\kappa) = m(\boldsymbol{x};\kappa,\beta)$ si ha che

$$g(\kappa)/g(\kappa+1) = \frac{\Gamma(\kappa+1)}{\Gamma(\kappa)}\frac{\Gamma(n+1+\kappa)}{\Gamma(n+1+\kappa+1)}\frac{\beta^\kappa}{\beta^{\kappa+1}}\frac{(\beta+x_n)^{n+1+\kappa+1}}{(\beta+x_n)^{n+1+\kappa}}.$$

Ricordando che, per ogni $\alpha > 0$, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, l'espressione qui sopra si semplifica in

$$g(\kappa)/g(\kappa+1) = \frac{\kappa}{n+1+\kappa}\frac{\beta+x_n}{\beta}.$$

Quindi la diseguaglianza $g(\kappa)/g(\kappa+1) < 1$ è verificata per tutti i $\kappa < \beta\frac{n+1}{x_n} \simeq 9.1448$. Quindi la funzione $g(\kappa)$ è crescente per $\kappa = 1, 2, \ldots, 9$ ed è decrescente per $\kappa = 10, 11, \ldots$. Inoltre è facile verificare che $g(9) < g(10)$. In conclusione $\kappa = 10$ è il punto di massimo di $g(\kappa)$.

7. Sia X_m^{new} la distanza percorsa dalla tartaruga in mezz'ora in una futura gara; poiché ci sono 3 intervalli di 10 min in mezz'ora, qui $m = 2$. Si tratta di calcolare:

$$\mathbb{P}(X_2^{new} > 4|\boldsymbol{x}) = \int_0^{+\infty} P(X_2^{new} > 4|\theta)\pi_3(\theta|\boldsymbol{x})d\theta,$$

dove $\pi_3(\cdot|\boldsymbol{x}) = gamma(\kappa_n, \beta_n)$ con $\kappa_n = \kappa + n + 1 = 22$ e $\beta_n = \beta + x_n = 41.62$. La legge condizionale di X_2^{new}, dato θ, è $gamma(3, \theta)$ e quindi, grazie al suggerimento:

$$\mathbb{P}(X_2^{new} > 4|\theta) = \mathrm{e}^{-4\theta} \sum_{i=0}^{2} \frac{(4\theta)^i}{i!}.$$

Pertanto

$$\begin{aligned}
\mathbb{P}(X_2^{new} > 4|\boldsymbol{x}) &= \int_0^{+\infty} \mathrm{e}^{-4\theta} \sum_{i=0}^{2} \frac{(4\theta)^i}{i!} \frac{\beta_n^{\kappa_n}}{\Gamma(\kappa_n)} \theta^{\kappa_n - 1} \mathrm{e}^{-\beta_n \theta} d\theta \\
&= \sum_{i=0}^{2} \frac{4^i}{i!} \frac{\beta_n^{\kappa_n}}{\Gamma(\kappa_n)} \int_0^{+\infty} \theta^{i+\kappa_n-1} \mathrm{e}^{-\theta(\beta_n+4)} d\theta = \sum_{i=0}^{2} \frac{4^i}{i!} \frac{\Gamma(i+\kappa_n)}{\Gamma(\kappa_n)} \frac{\beta_n^{\kappa_n}}{(\beta_n+4)^{i+\kappa_n}} \\
&= \left(\frac{\beta_n}{\beta_n+4}\right)^{\kappa_n} + 4\kappa_n \frac{\beta_n^{\kappa_n}}{(\beta_n+4)^{\kappa_n+1}} + 8\kappa_n(\kappa_n+1) \frac{\beta_n^{\kappa_n}}{(\beta_n+4)^{\kappa_n+2}} \\
&\simeq 0.1328 + 0.2562 + 0.2583 = 0.6473.
\end{aligned}$$

∎

03/07/2020 – Ex. 1

1. $X_1,\dots,X_n | \mu,\lambda \sim f(x;\lambda,\mu=2) = \frac{1}{2\lambda} e^{-\frac{1}{\lambda}|x-\mu|} = \frac{1}{2\lambda} e^{-\frac{1}{\lambda}|x-2|}$

Likelihood:

$$L(\underline{x};\mu=2,\lambda) = \prod_{i=1}^n f(x;\lambda,\mu=2) = \prod_{i=1}^n \frac{1}{2\lambda} e^{-\frac{1}{\lambda}|x_i-\mu|} = \left(\frac{1}{2\lambda}\right)^n e^{-\frac{1}{\lambda}\sum_{i=1}^n |x_i-\mu|}$$

$$= 2^{-n}\left(\frac{1}{\lambda}\right)^n e^{-\frac{1}{\lambda}\left(\sum_{i=1}^n |x_i-2|\right)}$$

$\Longrightarrow$ $\lambda \sim$ Inverse-Gamma(α,β) : $\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} e^{-\frac{\beta}{\lambda}}$

in this way $\mathbb{E}[\lambda] = \frac{\beta}{\alpha-1}$

Posterior:

$$\pi(\lambda|\underline{X}) \propto L(\underline{x},\lambda)\,\pi(\lambda) = \left[2^{-n}\lambda^{-n} e^{-\frac{1}{\lambda}\left(\sum_{i=1}^n |x_i-2|\right)}\right]\left[\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} e^{-\frac{\beta}{\lambda}}\right] \mathbb{1}\dots$$

$$\propto \left(2^{-n}\frac{\beta^\alpha}{\Gamma(\alpha)}\right) \cdot \lambda^{-(\alpha+n)-1} e^{-\frac{1}{\lambda}\left[\sum_{i=1}^n |x_i-2| + \beta\right]} \mathbb{1}\dots$$

$$\pi(\lambda|\underline{X}) \propto \text{IGamma}\left(\alpha+n,\ \beta+\sum_{i=1}^n |x_i-2|\right)$$

2. $\underline{x}^{old} = [1.7,\ 1.9,\ 2.05,\ 2.15,\ 2.2]^T$

$$\frac{\partial}{\partial\lambda}\log L(\underline{x},\lambda) = \frac{\partial}{\partial\lambda}\left[\log\left(\left(\frac{1}{\lambda}\right)^n\right) - \frac{1}{\lambda}\sum_{i=1}^n |x_i-2|\right] = \frac{\partial}{\partial\lambda}\left[-n\log(\lambda) - \frac{1}{\lambda}\sum_{i=1}^n |x_i-2|\right]$$

$$= -\frac{n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^n |x_i-2| = 0 \Longrightarrow \hat{\lambda}_{MLE} = \frac{\sum_{i=1}^n |x_i-2|}{n}$$

$$\mathbb{E}[\lambda|\underline{X}] = \frac{\beta+\sum_{i=1}^n |x_i-2|}{\alpha+n-1} = \frac{\alpha-1}{\alpha+n-1}\underbrace{\frac{\beta}{\alpha-1}}_{\mathbb{E}[\lambda]} + \frac{n}{\alpha+n-1}\underbrace{\frac{\sum_{i=1}^n |x_i-2|}{n}}_{\hat{\lambda}_{MLE}}$$

$\Longrightarrow$ $\begin{cases} \alpha-1 \text{ plays the role of sample-size,} \\ \beta \text{ plays the role of } \sum_{i=1}^n |x_i-2| \end{cases}$

$\Longrightarrow$ $\begin{cases} \alpha-1 = 5 \\ \beta = \sum_{i=1}^5 |x_i-2| \end{cases} \Longrightarrow \begin{cases} \alpha = 6 \\ \beta = 4/5 = 0.8 \end{cases}$

$\Longrightarrow$ $\pi_1(\lambda) \stackrel{d}{=}$ Inverse-gamma $(6,\ 0.8)$

3. $\underline{x}$: $n = 75$, $\sum_{i=1}^n |x_i-2| = 19.7$

$\Longrightarrow$ $\pi_1(\lambda|\underline{x}) \stackrel{d}{=}$ Inverse-Gamma $(6+75,\ 0.8+19.7) = (81,\ \frac{41}{2}) = (81, 20.5)$

$$\mathbb{E}[\lambda|\underline{X}] = \frac{20.5}{80} = 0.256$$

$$\text{var}(\lambda|\underline{X}) = \frac{(20.5)^2}{80^2 \cdot 79} = 0,00083$$

$\lambda|\underline{x}$ s.t. $\begin{matrix}\mathbb{E}[\lambda|\underline{x}] = 0.256 \\ \text{Var}(\lambda|\underline{x}) = 0,00083\end{matrix} \Big\} \Longrightarrow \lambda|\underline{x} \approx N(0.256,\ 0.0083)$

$$CI_{0.95} = \mathbb{E}[\lambda|\underline{x}] \pm z_{0.975}\sqrt{\text{var}(\lambda|\underline{x})} = [0.200\ ;\ 0.312]$$

4. Jeffreys: $I(\lambda) = \mathbb{E}\left[-\frac{\partial^2}{\partial\lambda^2}\log f(x,\lambda) \middle| \lambda\right]$

$$\frac{\partial^2}{\partial\lambda^2}\log f(x,\lambda) = \frac{\partial^2}{\partial\lambda^2}\left[-\log(2\lambda) - \frac{1}{\lambda}|x-2|\right] = \frac{\partial}{\partial\lambda}\left[-\frac{\cancel{2}}{\cancel{2}\lambda} + \frac{1}{\lambda^2}|x-2|\right] = \frac{1}{\lambda^2} - \frac{2|x-2|}{\lambda^3}$$

$$\mathbb{E}\left[\frac{1}{\lambda^2} - \frac{2|X-2|}{\lambda^3}\middle|\lambda\right] = \frac{1}{\lambda^2} - \frac{2}{\lambda^3}\mathbb{E}[|X-2|] = \frac{1}{\lambda^2} - \frac{2}{\lambda^3}\cdot\lambda = \frac{1}{\lambda^2} - \frac{2}{\lambda^2} = -\frac{1}{\lambda^2}$$

$$\Rightarrow I(\lambda) = \frac{1}{\lambda^2} \Rightarrow \pi_J(\lambda) \propto \sqrt{I(\lambda)} = \lambda^{-1}\mathbb{1}_{(0,+\infty)}(\lambda)$$

$$\pi(\lambda|\underline{x}) \propto L(\underline{x},\lambda)\,\pi_J(\lambda) \propto \lambda^{-n} e^{-\frac{1}{\lambda}\sum_{i=1}^n |x_i-2|}\,\lambda^{-1}\,\mathbb{1}_{(0,+\infty)}$$

$$\pi(\lambda|\underline{x}) \overset{d}{=} \text{Inv-Gamma}\left(n, \textstyle\sum_{i=1}^n |x_i-2|\right) = (75, 19.7)$$

5. $$f(x_{n+1}|\underline{x}) = \int_\Lambda f(x_{n+1}|\lambda)\,\pi(\lambda|\underline{x})\,d\lambda$$

$$= \int_0^\infty \frac{1}{2\lambda} e^{-\frac{1}{\lambda}|x_{n+1}-2|} \cdot \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\left(\frac{1}{\lambda}\right)^{-(\alpha_n+1)} e^{-\frac{\beta_n}{\lambda}}\,d\lambda$$

$$= \int_0^\infty \left(\frac{1}{2}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\right)\lambda^{-(\alpha_n+1)} e^{-\frac{1}{\lambda}[\beta_n + |x_{n+1}-2|]}\,d\lambda \qquad \text{Inv-}\mathcal{G}(\alpha_n, \beta_n)$$

$$= \left[\frac{1}{2}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\right]\frac{\Gamma(\alpha_n+1)}{(\beta_n+|x_{n+1}-2|)^{\alpha_n+1}} \qquad \alpha_n = 81,\quad \beta_n = 20.5$$

$$f(2.5|\underline{x}) = 0.274$$

16/06/2020 – Ex. 1

1. Likelihood: $L(\underline{x},\theta) = \prod_{i=1}^n f(x_i,\theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbb{1}_{(0,\theta)}(x_i) = \left(\frac{1}{\theta}\right)^n \mathbb{1}_{[0,\theta]}(\underline{x})$

$$\stackrel{\theta}{=} \frac{1}{\theta^n} \mathbb{1}_{[x_{max},+\infty)}(\theta)$$

$$\Longrightarrow \pi(\theta) \stackrel{d}{=} \pi(\theta; m, k) :\stackrel{d}{=} \frac{k m^k}{\theta^{k+1}} \mathbb{1}_{[m,+\infty)}(\theta)$$

$$\pi(\theta|\underline{x}) \propto L(\underline{x},\theta)\,\pi(\theta) = L(\underline{x},\theta)\,\pi(\theta; m, k) = \frac{1}{\theta^n} \mathbb{1}_{[x_{max},+\infty)}(\theta) \frac{k m^k}{\theta^{k+1}} \mathbb{1}_{[m,+\infty)}(\theta)$$

$$\pi(\theta|\underline{x}) \propto \frac{k m^k}{\theta^{k+n+1}} \mathbb{1}_{[\max\{x_{max}, m\},+\infty)}(\theta)$$

$$\propto \frac{(k+n)\, m_*^{(k+n)}}{\theta^{k+n+1}} \mathbb{1}_{[\,\cdot\,,+\infty)}(\theta)$$

$\Longrightarrow$ Pareto distr. with $k_1 = k+n$, $m_1 = \max\{x_{max}, m\}$

2. $\underline{x} = [4.2,\ 5.6,\ 6.1,\ 2.5,\ 3.4]$, $n=5$, $m=3$, $k=5$

Posterior hyperparams:

$m_1 = \max\{x_{max}, 3\} = \max\{6.1, 3\} = 6.1$

$k_1 = 5+5 = 10$

$\Longrightarrow$ Pareto $(k_1 = 10,\ m_1 = 6.1)$

Posterior point est. under quadratic loss function $\Longrightarrow$ post. expectation

$$\mathbb{E}[\theta|\underline{X}] = \int_{m_1}^{+\infty} \frac{k_1 m_1^{k_1}}{\theta^{k_1+1}} \cdot \theta \, d\theta = \int_{m_1}^{+\infty} k_1 m_1^{k_1} \cdot \theta^{-k_1} \, d\theta$$

$$= k_1 m_1^{k_1} \left[\frac{\theta^{-k_1+1}}{-k_1+1}\right]_{m_1}^{+\infty} = \frac{k_1 m_1^{k_1}}{1-k_1} \left[\frac{1}{\theta^{k_1-1}}\right]_{m_1}^{+\infty} = \frac{10\cdot(6.1)^{10}}{-9} \left[\frac{1}{\theta^9}\right]_{m_1}^{+\infty}$$

$$= \frac{10\cdot(6.1)^{10}}{9} \cdot \frac{1}{(6.1)^9} = \frac{10\cdot 6.1}{9} = \frac{61}{9} = 6{,}7778$$

3. $\begin{cases} H_0: \theta = 7 \\ H_1: \theta \neq 7 \end{cases}$

$m=3$, $k=5$:

$$\pi(\theta) = \frac{k m^k}{\theta^{k+1}} \mathbb{1}_{[m,+\infty)}(\theta)$$

Prior: $\pi_1(\theta) = \pi$

$$BF_{01} = \frac{\prod_{i=1}^n f(x_i; \theta = \theta_0)}{m_1(\underline{x})}$$

$$m_1(\underline{x}) = \int_\Theta f(\underline{x}|\theta)\,\pi_1(\theta)\,d\theta = \int_\Theta \prod_{i=1}^n f(x_i|\theta)\,\pi_1(\theta)\,d\theta$$

$$= \int_\Theta \frac{1}{\theta^n} \mathbb{1}_{[x_{max},+\infty)}(\theta) \frac{k m^k}{\theta^{k+1}} \mathbb{1}_{[m,+\infty)}(\theta)\,d\theta$$

$$= \int_\Theta \frac{k m^k}{\theta^{(k+n)+1}} \mathbb{1}_{[\max\{x_{max}, m\},+\infty)}(\theta)\,d\theta$$

$$= k m^k \cdot \frac{1}{(k+n)(\max\{x_{max}, m\})^{k+n}} = \frac{k m^k}{k_1 m_1^{k_1}} = \frac{5\cdot 3^5}{10\cdot 6.1^{10}}$$

$$BF_{01} = \frac{\left(\frac{1}{\theta_0}\right)^n \mathbb{1}_{[x_{max},+\infty)}(\theta)}{m_1(\underline{x})} = \frac{\left(\frac{1}{7}\right)^5 \cdot 10\cdot(6{,}1)^{10}}{5\cdot 3^5} = 34.93 \Longrightarrow H_0$$

$2\log BF_{01} = 7.1068$

4. $\pi(\theta) \stackrel{d}{=} \frac{k m^k}{\theta^{k+1}} \mathbb{1}_{[m,+\infty)}(\theta)$ $k=5$, $m=3$

$\mathbb{P}(X_6 > 6.5 | \underline{x}) = \int_{\Theta} \mathbb{P}(X_6 > 6.5 | \theta) \quad \pi(\theta|\underline{x})\, d\theta$

$\mathbb{P}(X_6 > 6.5 | \theta) = \int_{6.5}^{+\infty} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)\, dx = \frac{1}{\theta} \mathbb{1}_{\{\theta \in [6.5,+\infty)\}} \int_{6.5}^{\theta} dx$

$= \frac{1}{\theta} \mathbb{1}_{[6.5,+\infty)}(\theta)\ (\theta - 6.5) = \left(1 - \frac{6.5}{\theta}\right) \mathbb{1}_{[6.5,+\infty)}(\theta)$

$\mathbb{P}(X_6 > 6.5 | \underline{x}) = \int_{\Theta} \left(1 - \frac{6.5}{\theta}\right) \mathbb{1}_{[6.5,+\infty)}(\theta) \frac{k_1 m_1^{k_1}}{\theta^{k_1+1}} \mathbb{1}_{[m_1,+\infty)}(\theta)\, d\theta$

$= \int_{6.5}^{+\infty} \left(1 - \frac{6.5}{\theta}\right) k_1 m_1^{k_1} \frac{1}{\theta^{k_1+1}}\, d\theta = [\,..\,] = \frac{1}{11}\left(\frac{6.1}{6.5}\right)^{10}$

16/06/2020 – Ex. 2

$X_1, .., X_n | \theta \sim Be(\theta)$ $\theta \in (0,1)$

Discrete prior distribution for θ:

$\theta = \bar\theta$	0.2	0.4	0.6	0.8
$\mathbb{P}(\theta = \bar\theta)$	0.1	0.2	0.4	0.3

1. $\underline{x} = [0, .., 0, 1, .., 1]$ $\#0 = 9$, $\#1 = 6$

likelihood:

$L(\underline{x}, \theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^m (1-\theta)^{n-m} \mathbb{1}_{(0,1)}(\theta)$

$m = 6$

$n - m = 15 - 6 = 9$

θ	$\pi(\theta)$	$L(\theta; \underline{x})\pi(\theta)$	$\pi(\theta\|\underline{x})$
0.2	0.1	$(0,2)^6 (1-0,2)^9 \cdot 0,1$	0,0612
0.4	0.2	.	0,5887
0.6	0.4	.	0,3483
0.8	0.3	.	0,0028

colonna precedente normalizzata

2. Quadratic loss functions $\Rightarrow$ posterior exp.

$\mathbb{E}[\theta|\underline{x}] = \sum_{i=1}^n \theta \cdot \pi(\theta|\underline{x}) = [\,..\,] = 0.45894$

loss function 0-1 $\Rightarrow$ mode $\Longrightarrow$ $\arg\max \pi(\theta|\underline{x}) = 0.4$

3. $\begin{cases} H_0 : \theta \in \{0.2, 0.4\} \\ H_1 : \theta \in \{0.6, 0.8\} \end{cases}$

$BF_{01} = \frac{\mathbb{P}(\theta \in \{0,2,0.4\} | \underline{x})}{\mathbb{P}(\theta \in \{0.6, 0.8\} | \underline{x})} \cdot \frac{\mathbb{P}(\theta \in \{0.6, 0.8\})}{\mathbb{P}(\theta \in \{0.2, 0.4\})}$

$= \frac{\pi(0.2|\underline{x}) + \pi(0.4|\underline{x})}{\pi(0.6|\underline{x}) + \pi(0.8|\underline{x})} \cdot \frac{\pi(0.6) + \pi(0.8)}{\pi(0.2) + \pi(0.4)}$

$= \frac{(0.0612 + 0.5887)}{(0.3483 + 0.0028)} \cdot \frac{0.4 + 0.3}{0.1 + 0.2}$

$= 4.319$

$2 \log BF_{01} = 2.926 \Longrightarrow$ weak for H_0

4. Predictive:

$p(x_{n+1} | \underline{x}) = \sum_{\Theta} p(x_{n+1}|\theta)\pi(\theta|\underline{x})$

$= \sum_{\Theta} \theta^{x_{n+1}} (1-\theta)^{1-x_{n+1}} \pi(\theta|\underline{x})$

$= \begin{cases} 1 & \text{with prob} = 0.4589 \\ 0 & \text{with prob} = 0.5411 \end{cases}$

04/02/2020 – Ex. 1

1. $\underline{x} = (x_1, \dots, x_n, x_1^*, \dots, x_m^*)$

$\underline{\delta} = (1, \dots, 1, 0, \dots, 0)$

Survival function i: $S_i(t) = 1 - F_i(t) = 1 - \int_{-\infty}^{t} f_i(x_i, \lambda)\, dx_i$

$S_i(t) = 1 - \int_{-\infty}^{t} \frac{2^\lambda \lambda}{x^{\lambda+1}} \mathbb{1}_{[2,\infty)}(x)\, dx$

$= 1 - \left[\int_2^t \frac{2^\lambda \lambda}{x^{\lambda+1}} dx\right] \mathbb{1}_{[2,\infty)}(t) = 1 + 2^\lambda \cancel{\lambda} \left[\frac{x^{-\lambda}}{-\cancel{\lambda}}\right]_2^t \mathbb{1}_{[2,\infty)}(t)$

$= 1 + 2^\lambda \left[t^{-\lambda} - 2^{-\lambda}\right] \mathbb{1}_{[2,\infty)}(t) = 1 + 2^\lambda \left(\frac{1}{t^\lambda} - \frac{1}{2^\lambda}\right) \mathbb{1}_{[2,\infty)}(t)$

$= 1 + \left(\frac{2^\lambda}{t^\lambda} - 1\right) \mathbb{1}_{[2,\infty)}(t)$

2. Likelihood: $L(\theta|\underline{x}) \propto \prod_{i=1}^{n} \left[\left(f_i(x_i|\theta)\right)^{\delta_i} \left(S_i(x_i|\theta)\right)^{1-\delta_i} \right]$

$L(\theta|\underline{x}) \propto \prod_{i=1}^{n+m} \left(\frac{2^\lambda \cdot \lambda}{x_i^{\lambda+1}} \mathbb{1}_{[2,+\infty)}(x_i) \right)^{\delta_i} \left(\frac{2^\lambda}{x_i^\lambda}\right)^{1-\delta_i}$

$\propto \prod_{i=1}^{n+m} \left(\frac{2^\lambda \cdot \lambda}{x_i^{\lambda+1}}\right)^{\delta_i} \left(\frac{2^\lambda}{x_i^\lambda}\right)^{1-\delta_i} = \lambda^n 2^{(n+m)\lambda} \left[\prod_{i=1}^{n} \frac{1}{x_i^{\lambda+1}}\right] \left[\prod_{i=1}^{m} \frac{1}{x_i^{*\lambda}}\right]$

$\propto \lambda^n e^{-(\lambda+1)\sum_{i=1}^n \log(x_i)} e^{-\lambda \sum_{i=1}^m \log(x_i^*)} e^{(n+m)\log(2)\lambda}$

$\propto \lambda^n e^{-\lambda\left[\sum_{i=1}^n \log\left(\frac{x_i}{2}\right) + \sum_{i=1}^m \log\left(\frac{x_i^*}{2}\right)\right]}$ $\quad S_x \; ; \; S_{x*}$

$\longrightarrow$ Conj: $\lambda \sim \text{Gamma}(\alpha, \beta)$

$\pi(\lambda|\underline{x}) \propto L(\lambda, \underline{x})\, \pi(\lambda) \propto \lambda^n e^{-\lambda\left[\sum_{i=1}^n \dots + \sum_{i=1}^m \dots\right]} \lambda^{\alpha-1} e^{-\lambda\beta}$

$\propto \lambda^{n+\alpha-1} e^{-\lambda\left[\beta + \sum_{i=1}^n + \sum_{i=1}^m\right]}$

$\longrightarrow \pi(\lambda|\underline{x}) \overset{d}{=} \text{Gamma}\left(\alpha + n, \beta + \sum_{i=1}^n \log\left(\frac{x_i}{2}\right) + \sum_{i=1}^m \log\left(\frac{x_i^*}{2}\right)\right)$

3. $\underline{z} = (4.2, 4.6, 5.1, 5.2, 5.8)$ $\quad n = 5, \; m = 0$

$\alpha = 5$

$m(\underline{x}) = \int_\Lambda L(\lambda; \underline{x})\, \pi(\lambda)\, d\lambda = \int_\Lambda \lambda^n e^{-\lambda(S_x + S_{x*})} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\, d\lambda$

$[\dots] = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(n+\alpha)}{(S_x + S_{x*} + \beta)^{n+\alpha}}$

$S_{x*} = 0$

$\alpha = 5$

$\frac{\partial}{\partial\lambda} m(\underline{x}) = 0 \Rightarrow \frac{\partial}{\partial\lambda}\left(\frac{\beta^\alpha}{(S_x + S_{x*} + \beta)^{n+\alpha}}\right) = 0 \longrightarrow \beta = S_x = 4.531$

4. $n = 84$ $\quad \sum_{i=1}^n \log(x_i) = 110.355$

$m = 16$ $\quad \sum_{i=1}^m \log(x_i^*) = 20.371$

$S_x = \sum_{i=1}^n \log\left(\frac{x_i}{2}\right) = \sum_{i=1}^n \left(\log(x_i) - \log(2)\right) = \left[\sum_{i=1}^n \log(x_i)\right] - n \cdot \log(2)$

$S_{x*} = \quad \dots \quad \dots \quad = \left[\sum_{i=1}^m \log(x_i)\right] - m \log(2)$

$\longrightarrow \pi(\lambda|\underline{x}) \overset{d}{=} \text{Gamma}\left(89, \; 4.531 + (110.355 - 84 \cdot \log(2)) + (20.\cancel{371} - 16 \log(2))\right)$

$= \text{Gamma}(89, \; 65.943)$

$\longrightarrow \mathbb{E}[\lambda|\underline{X}] = \frac{89}{65.943} = 1.350$

5. $I(\lambda) = -\mathbb{E}\left[\frac{\partial^2}{\partial\lambda^2} \log f(x,\lambda) \mid \lambda\right]$

$$\frac{\partial^2}{\partial\lambda^2}\left(\log\left(2^\lambda \cdot \lambda \cdot x^{-(\lambda+1)}\right)\right) = \frac{\partial^2}{\partial\lambda^2}\left[\lambda\log(2) + \log(\lambda) - (\lambda+1)\log(x)\right]$$

$$= \frac{\partial}{\partial\lambda}\left[\log(2) + \frac{1}{\lambda} - \log(x)\right]$$

$$= -\frac{1}{\lambda^2} \Longrightarrow I(\lambda) = \frac{1}{\lambda^2} \Rightarrow \pi_J(\lambda) \propto \sqrt{1/\lambda^2}$$

$$\propto \lambda^{-1}\mathbb{1}_{(0,\infty)}(\lambda)$$

$$\pi(\lambda|\underline{x}) \propto L(\lambda;\underline{x})\,\pi_J(\lambda)$$

$$\propto \lambda^n e^{-\lambda(S_x + S_{x^*})}\frac{1}{\lambda}\mathbb{1}_{(0,+\infty)}(\lambda)$$

$$\propto \lambda^{n-1} e^{-\lambda(S_x+S_{x^*})} = \lambda^{83} e^{-\lambda(52.131 + 9.281)}$$

6. $\pi(\lambda|\underline{x}) \overset{d}{=} \text{Gamma}(89, 65.943)$

$\pi(\lambda) \overset{d}{=} \text{Gamma}(5, 4.531)$

$$\begin{cases} H_0: \lambda \le 1.6 \\ H_1: \lambda > 1.6 \end{cases}$$

$$BF_{01} = \frac{\mathbb{P}(\lambda \le 1.6|\underline{X})}{\mathbb{P}(\lambda > 1.6|\underline{X})} \cdot \frac{\mathbb{P}(\lambda > 1.6)}{\mathbb{P}(\lambda \le 1.6)}$$

$$\mathbb{P}(\lambda \le 1.6) = \mathbb{P}(2\lambda \cdot 4.531 < 1.6 \cdot 2 \cdot 4.531) = \mathbb{P}(\chi^2_{10} < 14.5) = 0.849$$

Posterior: $\mathbb{E}[\lambda|\underline{x}] = \frac{89}{65.943} = 1.350$

$$\text{var}(\lambda|\underline{x}) = \frac{89}{(65.943)^2} = 0.0204$$

$\Rightarrow$ $\pi(\lambda|\underline{x}) \approx N(1.350, 0.020)$

$\Rightarrow$ $\mathbb{P}(\lambda \le 1.6|\underline{X}) = \mathbb{P}\left(\frac{\lambda - 1.350}{\sqrt{0.02}} \le \frac{1.6-1.350}{\sqrt{0.02}}\right) = \Phi(1.768) = 0.961$

$\Rightarrow$ $BF_{01} = \dots = 4.382 \Rightarrow 2\log BF_{01} = 2.955 \Rightarrow H_0$

7. $\pi(\lambda|\underline{x}) \overset{d}{=} \text{Gamma}(89, 65.943)$

$\pi(\lambda) \overset{d}{=} \text{Gamma}(5, 4.531)$

X_{new}: pred. prob. that : $\mathbb{P}(X_{new} > 5|\underline{X})$:

$$\mathbb{P}(X_{new} > 5|\underline{X}) = 1 - \mathbb{P}(X_{new} \le 5|\underline{X}) = 1 - \int \mathbb{P}(X_{new} \le 5|\ )\,\pi(\lambda|\underline{X})\,d\lambda$$

$$= 1 - \left[\int_\Lambda \left[\int_{-\infty}^{5} \frac{2^\lambda \cdot \lambda}{x^{\lambda+1}}\mathbb{1}_{[2,\infty)}(x)\,dx\right]\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\lambda^{\alpha_n-1}e^{-\beta_n\lambda}\mathbb{1}_{(0,\infty)}(\lambda)\,d\lambda\right]$$

$$= [\dots] = 1 - \left(1 - \frac{\beta_n^{\alpha_n}}{(\beta_n - \log(2/5))^{\alpha_n}}\right)$$

$$= \frac{\beta_n^{\alpha_n}}{(\beta_n - \log(2/5))^{\alpha_n}} = 0{,}2928$$

14/01/2020 - Ex. 1

1. $X_1, \dots, X_n | \lambda \overset{iid}{\sim} \mathcal{E}(\lambda)$ $\quad X_i$ = waiting time

$Y_1, \dots, Y_n | \mu \overset{iid}{\sim} \mathcal{E}(\mu)$ $\quad Y_i$ = duration of the service

Likelihood:

$$L(\mu, \lambda; \underline{y}, \underline{x}) = \prod_{i=1}^n f(x_i, \lambda) f(y_i, \mu) = \prod_{i=1}^n \lambda e^{-\lambda x_i} \mathbb{1}_{(0,\infty)}(\lambda) \mu e^{-\mu y_i} \mathbb{1}_{(0,\infty)}(\mu)$$

$$= (\lambda\mu)^n e^{-\lambda \sum_{i=1}^n x_i - \mu \sum_{i=1}^n y_i} \mathbb{1} \dots$$

$\pi(\mu, \lambda) = \pi(\mu)\pi(\lambda)$: $\quad \pi(\lambda) \overset{d}{=} \text{gamma}(a_\lambda, b_\lambda)$

$\pi(\mu) \overset{d}{=} \text{gamma}(a_\mu, b_\mu)$

Posterior:

$$\pi(\mu, \lambda | \underline{y}, \underline{x}) \propto L(\mu, \lambda; \underline{y}, \underline{x}) \pi(\mu, \lambda)$$

$$\propto \left[(\lambda\mu)^n e^{-\lambda \sum_{i=1}^n x_i - \mu \sum_{i=1}^n y_i}\right]\left[\frac{b_\lambda^{a_\lambda}}{\Gamma(a_\lambda)} \lambda^{a_\lambda - 1} e^{-b_\lambda \lambda} \cdot \frac{b_\mu^{a_\mu}}{\Gamma(a_\mu)} \mu^{a_\mu - 1} e^{-b_\mu \mu}\right]$$

$$\propto \left(\lambda^{n + a_\lambda - 1} e^{-\lambda(S_x + b_\lambda)}\right)\left(\mu^{n + a_\mu - 1} e^{-\mu(S_y + b_\mu)}\right)$$

$\pi(\mu, \lambda | \underline{y}, \underline{x}) = \pi(\mu | \underline{y}) \pi(\lambda | \underline{x})$: $\quad \pi(\mu | \underline{y}) \overset{d}{=} \text{gamma}(a_\mu + n, S_y + b_\mu)$

$\pi(\lambda | \underline{x}) \overset{d}{=} \text{gamma}(a_\lambda + n, S_x + b_\lambda)$

2. ESP:

$$\frac{\partial}{\partial\lambda} \log L = \frac{\partial}{\partial\lambda}\left[n \log(\lambda) - \lambda S_x\right] = \frac{n}{\lambda} - S_x = 0 \Rightarrow \hat{\lambda}_{MLE} = \frac{n}{S_x}$$

$$\hat{\lambda}_{MLE} = \frac{n}{S_x}, \quad \hat{\mu}_{MLE} = \frac{n}{S_y}$$

$$\mathbb{E}[\lambda | \underline{x}] = \frac{a_\lambda + n}{S_x + b_\lambda} \longrightarrow \left(\mathbb{E}[\lambda|\underline{x}]\right)^{-1} = \frac{S_x + b_\lambda}{a_\lambda + n} = \frac{n}{a_\lambda + n}\frac{S_x}{n} + \frac{a_\lambda}{a_\lambda + n}\frac{b_\lambda}{a_\lambda}$$

$\Rightarrow \begin{cases} a_\lambda = a_\mu = \text{sample size} \\ b_\lambda = b_\mu = S_x \text{ or } S_y \end{cases}$

$\Rightarrow \begin{cases} a_\lambda = a_\mu = 8 \\ b_\lambda = 16.2 \\ b_\mu = 10.7 \end{cases} \Rightarrow \quad \lambda \sim \mathcal{G}(8, 16.2), \quad \mu \sim \mathcal{G}(8, 10.7)$

3. $n = 12$, $S_x = 22.6$, $S_y = 18.4$ $\Rightarrow$

$\lambda | \underline{x} \sim \mathcal{G}(20, 38.8)$ $\quad \mathbb{E}[\lambda|\underline{x}] = \frac{20}{38.8} = 0.5155$

$\mu | \underline{y} \sim \mathcal{G}(20, 29.1)$ $\quad \mathbb{E}[\mu|\underline{y}] = \frac{20}{29.1} = 0.6873$

4. $\begin{cases} H_0: \lambda = 0.5, \ \mu = 1 \\ H_1: \lambda \neq 0.5, \ \mu \neq 1 \end{cases}$ with prior of 2.

$$Bf_{01} = \frac{\prod_{i=1}^n f(x_i, \lambda = 0.5) f(y_i, \mu = 1)}{\int_\Theta f(\underline{x}|\lambda) f(\underline{y}|\mu) \pi(\lambda, \mu)\, d\lambda\, d\mu} = \frac{\left(\frac{1}{2}\right)^n e^{-\frac{1}{2}S_x - S_y}}{\left(\int \lambda^{a_\lambda + n - 1} e^{-\lambda(S_x + b_\lambda)} \frac{b_\lambda^{a_\lambda}}{\Gamma(a_\lambda)} d\lambda\right)(\dots(\mu))}$$

$$= [\dots] = \frac{\left(\frac{1}{2}\right)^n e^{-\frac{1}{2}S_x - S_y}}{\left[\frac{b_\lambda^{a_\lambda}}{\Gamma(a_\lambda)} \cdot \frac{\Gamma(a_\lambda + n)}{(S_x + b_\lambda)^{a_\lambda + n}}\right] \cdot \left[\frac{b_\mu^{a_\mu}}{\Gamma(a_\mu)} \cdot \frac{\Gamma(a_\mu + n)}{(S_y + b_\mu)^{a_\mu + n}}\right]} = \dots = -0.6122$$

$\Longrightarrow$ weak evidence in favor of H_1

5. $\mathbb{E}\left[\frac{\lambda}{\mu}\middle|\underline{x},\underline{y}\right] = \mathbb{E}[\lambda|\underline{x}]\,\mathbb{E}\left[\frac{1}{\mu}\middle|\underline{y}\right] = \left(\frac{a_\lambda+n}{b_\lambda+S_x}\right)\int \frac{1}{\mu}\cdot\frac{(b_\mu+S_y)^{a_\mu+n}}{\Gamma(a_\mu+n)}\mu^{a_\mu+n-1}e^{-(b_\mu+S_y)\mu}d\mu$

$$= \left(\frac{a_\lambda+n}{b_\lambda+S_x}\right)\frac{(b_\mu+S_y)^{a_\mu+n}}{\Gamma(a_\mu+n)}\frac{\Gamma(a_\mu+n-1)}{(b_\mu+S_y)^{a_\mu+n-1}}$$

$$= \frac{a_\lambda+n}{b_\lambda+S_x}\;\frac{(b_\mu+S_y)}{(a_\mu+n-1)}$$

$\lambda|\underline{x} \sim \Gamma(20, 38.8)$ $\qquad 38.8\;\lambda|\underline{x} \sim \Gamma(20,1) = \Gamma(\frac{40}{2},1) = \chi^2_{10}$

$\mu|\underline{y} \sim \Gamma(20, 29.1)$ $\qquad 29.1\;\;|\underline{y} \sim \Gamma(20,1) = \Gamma(\frac{40}{2},1) = \chi^2_{10}$

$$\frac{\frac{38.8\;\lambda|\underline{x}}{40}}{\frac{29.1\;\mu|\underline{x}}{40}} \sim F(40,40)$$

$$\mathbb{P}\left(\frac{\lambda}{\mu}<1\middle|\underline{x},\underline{y}\right) = \mathbb{P}\left(\frac{38.8}{40}\frac{40}{29.1}\frac{\lambda}{\mu} < \frac{38.8}{29.1}\middle|\underline{x},\underline{y}\right) = 0.8166$$

14/01/2020 - Ex.3

$\mathbb{P}(\sigma<t) = \frac{t}{\sigma_0} \Rightarrow F_{\sigma^2}(t) = \mathbb{P}(\sigma^2<t) = \mathbb{P}(\sigma<\sqrt{t}) = \frac{\sqrt{t}}{\sigma_0}$

$f_{\sigma^2}(t) = \frac{1}{2\sigma_0\sqrt{t}} \qquad \left(= \frac{d}{dt}F_{\sigma^2}(t)\right)$

$L(\underline{x},\mu,\sigma^2) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2}$

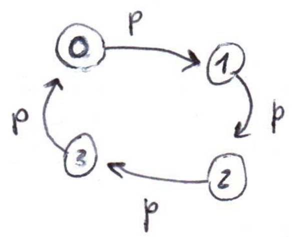
$\pi(\mu,\sigma^2|\underline{x}) \propto \left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2}\right]\cdot\left[\frac{1}{\sqrt{2\pi\tau_0^2}}e^{-\frac{1}{2\tau_0^2}(\mu-\eta_0)^2}\right]\cdot\left[\frac{1}{2\sigma_0\sigma}\right]$

$[\sigma^2|-] \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\frac{1}{2}} e^{-\frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\mu)^2} \sim \text{inv-}\Gamma\left(\frac{n-1}{2}, \frac{\sum_{i=1}^n(x_i-\mu)^2}{2}\right)$

$[\mu|-] \propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2-\frac{1}{2\tau_0^2}(\mu-\eta_0)^2} \sim N(\cdot,\cdot)$

13/06/2019 - Ex. 1

1.



X_n = state (at n-th), $X_n \in \{0,1,2,3\}$

n^+ = # clockwise

$n^- = n - n^+$

Likelihood: $\mathbb{P}(X_1 = x_1, .., X_n = x_n | X_0 = x_0, p)$ $\qquad x_0 = 0$

$\mathbb{P}(X_1 = x_1, .., X_n = x_n | X_0 = x_0, p) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, p) \cdots \mathbb{P}(X_1 = x_1 | X_0 = x_0, p)$

$$\mathbb{P}(X_j = x_j | X_{j-1} = x_{j-1}, p) = \begin{cases} p & x_{j-1} = x_j + 1 \\ 1-p & x_{j-1} = x_j - 1 \\ 0 & x_{j-1} \nexists \uparrow \end{cases}$$

$$\Rightarrow \mathbb{P}(\underline{X} = \underline{x} | X_0 = x_0, p) = \prod_{i=1}^{n} p^{\mathbb{1}\{x_{i-1} = x_i + 1\}} (1-p)^{\mathbb{1}\{x_{i-1} = x_i - 1\}}$$

$$= p^{n+} (1-p)^{n-} \mathbb{1}_{(0,1)}(p)$$

$$= p^{n+} (1-p)^{n-n+} \mathbb{1}_{(0,1)}(p)$$

2. Beta-Bernoulli

$p \sim Beta(\alpha, \beta)$

$$\pi(p|\underline{x}) \propto L(p, \underline{x})\, \pi(p) = p^{n^+}(1-p)^{n-n^+} \mathbb{1}_{(0,1)}(p)\; p^{\alpha-1}(1-p)^{\beta-1} \mathbb{1}_{(0,1)}(p)$$

$$= p^{n^+ + \alpha - 1} (1-p)^{n - n^+ + \beta - 1} \mathbb{1}_{(0,1)}(p)$$

$$\pi(p|\underline{x}) \sim Beta(\alpha + n^+, \beta + n - n^+)$$

3. $$\mathbb{E}[p|\underline{X}] = \frac{\alpha + n^+}{\alpha + \beta + n} = \frac{n}{\alpha+\beta+n}\frac{n^+}{n} + \frac{\alpha+\beta}{\alpha+\beta+n}\frac{\alpha}{\alpha+\beta}$$

$$\left(\frac{\partial}{\partial p} \log L = \frac{\partial}{\partial p}\left(n^+ \log(p) + (n-n^+)\log(1-p)\right) = \frac{n^+}{p} - \frac{n-n^+}{1-p} = 0 : \quad \hat{p}_{MLE} = \frac{n^+}{n} \right)$$

$\Rightarrow$ $\alpha + \beta$ plays the role of the size (n)

α plays the role of n^+

$\Rightarrow$ $\alpha + \beta = 10$, $\alpha = 3$ $\Rightarrow$ $\alpha = 3$, $\beta = 7$ $\Longrightarrow$ $\boxed{\pi_1(p) \overset{d}{=} Beta(3,7)}$

4. $$\pi_1(p|\underline{x}) \overset{d}{=} Beta(3 + n^+, 7 + n^-) = Beta(3+6, 7+9) = Beta(9, 16)$$

$$\mathbb{E}[p|\underline{x}] = \frac{9}{9+16} = 0.36 = \frac{9}{25} \qquad Var(p|\underline{x}) = \frac{9 \cdot 16}{(9+16+1)(9+16)^2} = 0.0089 = \frac{72}{8125}$$

5. $$\mathbb{P}(X_1 - X_0 = 1 | X_0 = 0) = \mathbb{P}(X_0 \text{ salta in avanti})$$

$$\mathbb{P}(X_0 \text{ salta in avanti}) = \int_0^1 \mathbb{P}(X_0 \text{ salta in avanti} | p)\, \pi(p)\, dp$$

$$= \int_0^1 p \frac{1}{B(\alpha+\beta)} p^{\alpha-1}(1-p)^{\beta-1} dp$$

$$= [\ldots] = \frac{\alpha}{\alpha+\beta}$$

We impose $\frac{\alpha}{\alpha+\beta} = 0.4 \implies \beta = \frac{3}{2}\alpha$

α	β	α_n	β_n	$\mathbb{E}[p \mid \underline{x}]$	$Var(p \mid \underline{x})$
3	7	9	16	0.36	0.0089
0.5	0,75	6.5	9.75	0.4	0.0139
1	1.5	7	10.5	0.4	0.0130
2	2	8	12	0.4	0.0114

robust ✓

6. $\mathbb{P}(m=j \mid \underline{x}, x_0) = \frac{m(\underline{x} \mid x_0, m=j)\ \mathbb{P}(m=j)}{m(\underline{x} \mid x_0)}$

$$m(\underline{x} \mid x_0, m=j) = \int f(\underline{x} \mid x_0, m=j, p_j)\ \pi(p_j)\, dp_j$$

$$= \int_0^1 p^{n^+}(1-p)^{n^-} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp$$

$$= \dots$$

$$= \frac{B(\alpha_j + n^+,\ \beta_j + n^-)}{B(\alpha_j, \beta_j)}$$

$$m(\underline{x} \mid x_0) = \sum_{j=1}^{4} m(\underline{x} \mid x_0, m=j)\ \mathbb{P}(m=j)$$

$$\mathbb{P}(m=j \mid \underline{x}, x_0) \propto \begin{cases} -10.752 & m=1 \\ -11.556 & m=2 \\ -11.156 & m=3 \\ -10.827 & m=4 \end{cases}$$

$\longrightarrow$ wins

14/02/2019

1. $X_1, \dots, X_n | \theta \sim \mathcal{E}(\theta)$, $\theta \sim \pi(\theta)$ θ = failure rate

$\pi_J(\theta) \propto \sqrt{|I(\theta)|}$

$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(x,\theta) | \theta \right]$

$\frac{\partial^2}{\partial\theta^2} \log f(x,\theta) = \frac{\partial^2}{\partial\theta^2} \log(\theta e^{-\theta x}) = \frac{\partial^2}{\partial\theta^2}\left[\log\theta - \theta x\right] = \frac{\partial}{\partial\theta}\left[\frac{1}{\theta} - x\right] = -\frac{1}{\theta^2}$

$\pi_J(\theta) \propto \theta^{-1} \mathbb{1}_{(0,\infty)}(\theta)$

$L(\underline{x},\theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^n x_i}$

$\pi(\theta|\underline{x}) \propto \theta^n e^{-\theta\sum_{i=1}^n x_i} \theta^{-1} \mathbb{1}_{(0,\infty)}(\theta)$

$\propto \theta^{n-1} e^{-\theta\sum_{i=1}^n x_i} \mathbb{1}_{(0,\infty)}(\theta) \sim \text{Gamma}(n, \sum_{i=1}^n x_i)$

$\mathbb{E}[\theta|\underline{x}] = \frac{n}{\sum_{i=1}^n x_i} \approx 0,3855 \quad \left(\frac{10}{25.94}\right)$

2. $\pi_2(\theta) \overset{d}{=} \text{Gamma}(\alpha, \beta)$

$\pi_2(\theta|\underline{x}) \propto \theta^n e^{-\theta S_x} \frac{\alpha^\beta}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta}$

$\propto \theta^{n+\alpha-1} e^{-\theta(\beta+S_x)} \implies \pi_2(\theta|\underline{x}) \overset{d}{=} \Gamma(n+\alpha, \beta+S_x)$

3. $\mathbb{E}[\theta|\underline{x}] = \frac{n+\alpha}{\beta+S_x}$

$\frac{\partial}{\partial\theta}\log L = \frac{\partial}{\partial\theta}\left(n\log\theta - \theta S_x\right) = 0 = \frac{n}{\theta} - S_x \Rightarrow \hat{\theta}_{MLE} = \frac{n}{S_x}$

$\mathbb{E}[\theta|\underline{x}] = \left(\frac{\beta+S_x}{n+\alpha}\right)^{-1} = \left(\frac{\alpha}{n+\alpha}\frac{\beta}{\alpha} + \frac{n}{n+\alpha}\frac{S_x}{n}\right)^{-1}$

$\alpha \rightsquigarrow n$

$\beta \rightsquigarrow S_x$

4. $m_{X_1}(x_1) = \int f(x|\theta)\pi(\theta)\,d\theta = \int_0^\infty \theta e^{-\theta x} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\theta\beta}\,d\theta$

$= \int_0^\infty \theta^\alpha e^{-\theta(\beta+x)}\,d\theta \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{(\beta+x)^{\alpha+1}} = \frac{\alpha\beta^\alpha}{(\beta+x)^{\alpha+1}}$

$F_{X_1}(x) = \int_0^x m_{X_1}(t)\,dt = \int_0^x \frac{\alpha\beta^\alpha}{(\beta+t)^{\alpha+1}}\,dt = \alpha\beta^\alpha \left[\frac{(\beta+t)^{-\alpha}}{-\alpha}\right]_0^x$

$= -\beta^\alpha\left[(\beta+x)^{-\alpha} - \beta^{-\alpha}\right]$

$= \left(\frac{1}{\beta^\alpha} - \frac{1}{(\beta+x)^\alpha}\right)\beta^\alpha = 1 - \frac{\beta^\alpha}{(\beta+x)^\alpha}$

$F(x) = p \Rightarrow 1 - \frac{\beta^\alpha}{(\beta+x)^\alpha} = p \Rightarrow 1-p = \frac{\beta^\alpha}{(\beta+x)^\alpha} \Rightarrow (1-p)^{\frac{1}{\alpha}} = \frac{\beta}{\beta+x}$

$\Rightarrow \beta + x = \frac{\beta}{(1-p)^{1/\alpha}} \Rightarrow x = \frac{\beta}{(1-p)^{1/\alpha}} - \beta = Q(p)$

5. $\alpha = 10$

$Q(0.975) - Q(0.025) = 5.02 - 0.29 = 4.73 \implies \beta = 10.663$

$\implies \pi_2(\theta) \overset{d}{=} \Gamma(10, 10.663)$, $\pi_2(\theta|\underline{x}) \overset{d}{=} \Gamma(20, 36.603) \Rightarrow \mathbb{E}[\theta|\underline{X}] = 0.5464$

6. $\begin{cases} H_0: & \frac{1}{\theta} > 2.5 \\ H_1: & \frac{1}{\theta} \leq 2.5 \end{cases}$ π_2

$$BF_{01} = \frac{\mathbb{P}(\frac{1}{\theta} > 2.5 \mid \underline{x})}{1 - \mathbb{P}(\frac{1}{\theta} > 2.5 \mid \underline{x})} \cdot \frac{1 - \mathbb{P}(\frac{1}{\theta} > 2.5)}{\mathbb{P}(\frac{1}{\theta} > 2.5)}$$

$$\mathbb{P}\left(\theta < \frac{1}{2.5}\right) = \mathbb{P}\left(\theta < \frac{2}{5}\right) = 0.0123$$

$$\theta \mid \underline{x} \approx N(\mathbb{E}[\theta \mid \underline{x}], Var(\theta \mid \underline{x})) = N(0.5464, 0.0149)$$

$$\mathbb{P}\left(\theta < \frac{2}{5} \mid \underline{x}\right) = \Phi\left(\frac{0.4 - 0.5464}{\sqrt{0.0199}}\right) = 1 - \Phi(1.984) \sim 0.1152$$

$BF_{01} = \ldots \Rightarrow 2\log BF_{01} = 4 \Longrightarrow H_0$

7. The two full conditionals are:
   1. $\mathcal{L}(z_1, \ldots, z_k \mid data, \theta)$
   2. $\mathcal{L}(\theta \mid data, z_1, \ldots, z_k)$