TP 2 – AOS1

Priors
Corrigé

1 Introduction

The purpose of the practical session is to consider various sampling distributions $p_X(\cdot|\theta)$, and to study the effect of various posterior distributions $\pi_\theta(\cdot|\eta)$. This study will be formal (computation of the posterior distribution $\pi_\theta(\cdot|X,\alpha)$ and of the posterior predictive distribution $p_X(\cdot|\eta)$) as well as graphical (representation of the prior, sampling, and aforementioned distributions).

2 Conjugate priors

2.1 Discrete sampling distributions

Binomial sampling distribution, beta prior

1 Recall the expression of the beta distribution. What is its definition domain ? On which parameters does it depend ? What are the expectation, the mode, and the variance ? Write a script which plots the prior distribution given a set of chosen hyper-parameter values. Which particular distribution can be retrieved as a special case of the beta distribution ?

The pdf of the beta distribution is defined by

$$\pi_\theta(t|\alpha,\beta) = \text{beta}(t;\alpha,\beta) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}, \text{ with } \mathrm{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

and Γ is the Gamma function. It depends on two parameters, written here $\alpha > 0$ and $\beta > 0$. We have

$$\mathbb{E}[\theta] = \frac{\alpha}{\alpha+\beta}, \quad \text{Mode}(\theta) = \frac{\alpha-1}{\alpha+\beta-2} \text{ for } \alpha > 1, \beta > 1, \quad \text{Var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

The pdf of the beta distribution can be represented using the following code.

```
import numpy as np
import scipy.stats as spst
import matplotlib.pyplot as plt

distrib = spst.beta(a=0.5,b=0.5)

t = np.arange(start=0, stop=1, step=0.01)

fig1, ax1 = plt.subplots()
ax1.plot(t, distrib.pdf(t))
```

A special case, retrieved with $\alpha = \beta = 1$, is the uniform distribution on the $[0;1]$ interval.

② Consider now a random variable X following a binomial sampling distribution $\mathcal{B}(n, \theta)$, with n known and $\theta \sim \text{beta}(\alpha, \beta)$. Recall the expression of the posterior distribution of θ given x, α and β, and plot $\pi_\theta(\cdot)$, $L(\cdot|x)$ and $\pi_\theta(\cdot|x; \alpha, \beta)$.

The posterior distribution of θ is also a beta distribution:

$$L(\theta|x) = p_X(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x};$$
$$\pi_\theta(t|x; \alpha, \beta) \propto L(t|x)\pi_\theta(t|\alpha, \beta) = \binom{n}{x} \frac{1}{\mathrm{B}(\alpha, \beta)} t^{x+\alpha-1}(1-t)^{n-x+\beta-1}.$$

It is common to interpret the hyper-parameters α and β of the prior as a number of "pseudo-observations" of X. In the binomial setting, x stands for the number of successes of the underlying independently repeated Bernoulli experiment, and its complement $n - x$ as the number of failures. If we properly compute the normalizing constant of the posterior distribution, we obtain

$$\pi_\theta(t|x; \alpha, \beta) = \frac{L(t|x)\pi_\theta(t|\alpha, \beta)}{\int_0^1 L(t|x)\pi_\theta(t|\alpha, \beta)dt} = \frac{t^{x+\alpha-1}(1-t)^{n-x+\beta-1}}{\mathrm{B}(\alpha + x, \beta + n - x)};$$

we recognize here the beta distribution with parameters $\alpha + x$ and $\beta + n - x$: then, α and β can therefore be thought of as numbers of virtual successes and failures. This also shows that the beta prior is conjugate for the binomial distribution.
The prior, sampling and posterior distributions can be plot using the following code.

```
nbin = 10
alph, beta = (5, 1)

sampl_dist = spst.binom(n=nbin, p=0.2)
prior_dist = spst.beta(a=alph, b=beta)

x = sampl_dist.rvs()
post_dist = spst.beta(a=alph+x, b=beta+nbin-x)

t = np.arange(start=0.01, stop=1, step=0.01)
ll = [loglike((nbin,p), spst.binom, x) for p in t]

fig, axs = plt.subplots(1, 3, sharex=True, tight_layout=True)
axs[0].plot(t, prior_dist.logpdf(t))
axs[1].plot(t, ll)
axs[2].plot(t, post_dist.logpdf(t))
```

③ Let us now assume that we have previously observed $X = x$ positive outcomes out of n outcomes. What is the predictive distribution for the number X_0 of positive outcomes out of n_0 new experiments, given $X = x$ out of n, α and β ?

The predictive distribution for X_0 is

$$\begin{aligned}
p_{X_0}(x_0|x; \alpha, \beta) &= \int_0^1 p_{X_0}(x_0|t)\pi_\theta(t|x; \alpha, \beta)dt, \\
&= \int_0^1 \binom{n_0}{x_0} t^{x_0}(1-t)^{n_0-x_0} \mathrm{Beta}(\alpha + x, \beta + n - x)dt, \\
&= \binom{n_0}{x_0} \frac{1}{\mathrm{B}(\alpha + x, \beta + n - x)} \int_0^1 t^{\alpha+x+x_0-1}(1-t)^{\beta+n-x+n_0-x_0-1}dt, \\
&= \binom{n_0}{x_0} \frac{\mathrm{B}(\alpha + x + x_0, \beta + n - x + n_0 - x_0)}{\mathrm{B}(\alpha + x, \beta + n - x)}.
\end{aligned}$$

Poisson sampling distribution, gamma prior

④ Recall the expression of the gamma distribution, its definition domain, the parameters on which it depends. Recall its expectation, mode, and variance. Write a script which plots the prior distribution given a set of chosen hyper-parameter values.

The pdf of the gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ is

$$\pi_\theta(t|\alpha,\beta) = \text{gamma}(\text{t};\alpha,\beta) = \frac{\beta^\alpha t^{\alpha-1}\exp(-\beta t)}{\Gamma(\alpha)}, \text{ with } \Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}\exp(-t)dt,$$

(again, Γ is the Gamma function). Note that the gamma distribution can also be expressed using parameters α and $\eta = 1/\beta$; generally, η is referred to as *scale* parameter, and β as *rate* parameter. The formulation using the former is more frequent, but using the latter makes computing the posterior distribution more obvious; we shall therefore use it. We have

$$\mathbb{E}[\theta] = \frac{\alpha}{\beta}, \quad \text{Mode}(\theta) = \frac{\alpha-1}{\beta} \text{ for } \alpha > 1, \quad \text{Var}(\theta) = \frac{\alpha}{\beta^2}.$$

The pdf of the beta distribution can be represented using the following code.

```
distrib = spst.gamma(a=9,scale=1/2)

t = np.arange(start=0, stop=15, step=0.1)

fig1, ax1 = plt.subplots()
ax1.plot(t, distrib.pdf(t))
```

Note that Python uses the scale parametrization.

⑤ Consider a random variable $X \sim \mathcal{P}(\theta)$, with $\theta \sim \text{gamma}(\alpha,\beta)$. What is the posterior distribution of θ given an iid sample $x_1,\dots,x_n$, α and β ? Plot $\pi_\theta(\cdot)$, $L(\cdot|x_1,\dots,x_n)$ and $\pi_\theta(\cdot|x_1,\dots,x_n;\alpha,\beta)$, for various values of θ, α, β and n.

The posterior distribution of θ is also a gamma distribution. Indeed, assuming $x_i \geqslant 0$ for all $i = 1,\dots,n$,

$$L(\theta|x_1,\dots,x_n) = \prod_{i=1}^n \Pr(X = x_i|\theta) = \exp(-n\theta)\frac{\theta^{\sum_i x_i}}{\prod_i x_i!};$$

$$\pi_\theta(t|x_1,\dots,x_n;\alpha,\beta) = \frac{L(t|x_1,\dots,x_n)\pi_\theta(t|\alpha,\beta)}{\displaystyle\int_0^{+\infty} L(t|x_1,\dots,x_n)\pi_\theta(t|\alpha,\beta)} = \frac{\exp\left(-(n+\beta)t\right)\, t^{\alpha+\sum_i x_i - 1}}{\displaystyle\int_0^{+\infty} \exp\left(-(n+\beta)t\right)\, t^{\alpha+\sum_i x_i-1}dt}.$$

Remark that

$$\begin{aligned}\int_0^{+\infty} \exp\left(-(n+\beta)t\right)\, t^{\alpha+\sum_i x_i-1}dt &= \frac{\Gamma(\alpha+\sum_i x_i)}{(\beta+n)^{\alpha+\sum_i x_i}}\int_0^{+\infty} \text{gamma}(t;\alpha+\sum_i x_i,\beta+n)dt \\ &= \frac{\Gamma(\alpha+\sum_i x_i)}{(\beta+n)^{\alpha+\sum_i x_i}};\end{aligned}$$

Hence, the gamma prior is conjugate for the Poisson distribution:

$$\pi_\theta(t|x_1,\dots,x_n;\alpha,\beta) = \text{gamma}\left(t;\alpha+\sum_i x_i,\beta+n\right).$$

The prior, sampling and posterior distributions can be plot using the same code as above.

⑥ Assume that an iid sample $x_1,\dots,x_n$ of realizations of $X \sim \mathcal{P}(\theta)$ has been observed. Show that the predictive distribution of a new outcome x_0 given the sample, α and β is a negative binomial (or Pólya) distribution. At some point, you may want to make a change of integration variable, by replacing t with $z = (\beta+n+1)t$.

Assuming $x_0 \geqslant 0$, we integrate with respect to θ:

$$\begin{aligned}
p_{X_0}(x_0|x_1,\dots,x_n;\alpha,\beta) &= \int_0^{+\infty} p_{X_0}(x_0|t)\pi_\theta(t|x_1,\dots,x_n;\alpha,\beta)dt,\\
&= \frac{(\beta+n)^{\alpha+\sum_i x_i}}{\Gamma(\alpha+\sum_i x_i)\,x_0!}\int_0^{+\infty} \exp\left(-(n+\beta+1)t\right)\, t^{\alpha+\sum_i x_i+x_0-1}dt.
\end{aligned}$$

Let us now make the change of variable:

$$z=(\beta+n+1)t \quad \Leftrightarrow \quad t=(\beta+n+1)^{-1}z \quad \Rightarrow \quad dt=(\beta+n+1)^{-1}dz;$$

since the bounds of the integral are not modified by this change of variable, this yields

$$p_{X_0}(x_0|x_1,\dots,x_n;\alpha,\beta) = \frac{(\beta+n)^{\alpha+\sum_i x_i}}{\Gamma(\alpha+\sum_i x_i)\,x_0!}\int_0^{+\infty} \frac{\exp\left(-z\right)\, z^{\alpha+\sum_i x_i+x_0-1}}{(\beta+n+1)^{\alpha+\sum_i x_i+x_0}}dz.$$

The integral corresponds to a specific value for the Gamma function Γ: we finally obtain

$$\begin{aligned}
p_{X_0}(x_0|x_1,\dots,x_n;\alpha,\beta) &= \frac{\Gamma(\alpha+\sum_i x_i+x_0)}{\Gamma(\alpha+\sum_i x_i)\,x_0!}\,\frac{(\beta+n)^{\alpha+\sum_i x_i}}{(\beta+n+1)^{\alpha+\sum_i x_i+x_0}},\\
&= \frac{\Gamma(\alpha+\sum_i x_i+x_0)}{\Gamma(\alpha+\sum_i x_i)\,x_0!}\left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum_i x_i}\frac{1}{(\beta+n+1)^{x_0}},\\
&= \frac{\Gamma(\alpha+\sum_i x_i+x_0)}{\Gamma(\alpha+\sum_i x_i)\,x_0!}\left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum_i x_i}\left(1-\frac{\beta+n}{\beta+n+1}\right)^{x_0},
\end{aligned}$$

which corresponds to the pdf of the "generalized" negative binomial distribution, when the numbers of successes and trials are not necessarily integers. Thus, we have

$$X_0|x_1,\dots,x_n,\alpha,\beta \sim \text{Neg Bin}\left(\alpha+\sum_i x_i,\frac{\beta+n}{\beta+n+1}\right).$$

2.2 Continuous sampling distributions

Exponential sampling distribution, gamma prior

7 Consider a random variable $X \sim \mathcal{E}(\theta)$, and $\theta \sim \text{gamma}(\alpha,\beta)$. What is the posterior distribution of θ given an iid sample $x_1,\dots,x_n$, α and β ?

We have here

$$p_\theta(t;x_1,\dots,x_n;\alpha,\beta) = \frac{L(t;x_1,\dots,x_n)\pi_\theta(t;\alpha,\beta)}{\int_0^{+\infty} L(t;x_1,\dots,x_n)\pi_\theta(t;\alpha,\beta)dt} = \frac{t^{\alpha+n-1}\exp(-(\beta+\sum_i x_i)t)}{\int_0^{+\infty} t^{\alpha+n-1}\exp(-(\beta+\sum_i x_i)t)dt}.$$

Since

$$\int_0^{+\infty} t^{\alpha+n-1}\exp(-(\beta+\sum_i x_i)t)dt = \frac{\int_0^{+\infty} z^{\alpha+n-1}\exp(-z)dz}{(\beta+\sum_i x_i)^{\alpha+n}} = \frac{\Gamma(\alpha+n)}{(\beta+\sum_i x_i)^{\alpha+n}},$$

the posterior distribution of θ given the sample $x_1,\dots,x_n$, α and β is a gamma distribution:

$$\begin{aligned}
p_\theta(t;x_1,\dots,x_n;\alpha,\beta) &= \frac{(\beta+\sum_i x_i)^{\alpha+n}\,t^{\alpha+n-1}\exp(-(\beta+\sum_i x_i)t)}{\Gamma(\alpha+n)},\\
&= \text{gamma}\left(\alpha+n,\beta+\sum_i x_i\right).
\end{aligned}$$

8 Compute the predictive distribution for a new random variable $X_0 \sim \mathcal{E}(\theta)$, given $x_1,\dots,x_n$, α and β. We recall that $\Gamma(u+1)=u\,\Gamma(u)$.

Assuming $x_0 > 0$, integrating with respect to θ and using $z = (\beta + \sum x_i + x_0)t$ gives:

$$\begin{aligned} p_{X_0}(x_0|x_1,\ldots,x_n;\alpha,\beta) &= \int_0^{+\infty} t\,\exp(-tx_0)\frac{(\beta+\sum_i x_i)^{\alpha+n}\,t^{\alpha+n-1}\exp(-(\beta+\sum_i x_i)t)}{\Gamma(\alpha+n)}dt,\\ &= \frac{(\beta+\sum_i x_i)^{\alpha+n}}{\Gamma(\alpha+n)}\int_0^{+\infty} t^{\alpha+n}\exp(-(\beta+\sum_i x_i + x_0)t)dt;\\ &= \frac{(\beta+\sum_i x_i)^{\alpha+n}\int_0^{+\infty} z^{\alpha+n}\exp(-z)dz}{\Gamma(\alpha+n)\,(\beta+\sum_i x_i+x_0)^{\alpha+n+1}} = \frac{(\alpha+n)\,(\beta+\sum_i x_i)^{\alpha+n}}{(\beta+\sum_i x_i+x_0)^{\alpha+n+1}}. \end{aligned}$$

Normal sampling distribution, normal-gamma prior We now consider a Gaussian random variable $X \sim \mathcal{N}(\mu,\lambda^{-1})$, where the Gaussian distribution is parameterized using the expectation μ and the *precision* $\lambda = 1/(\sigma^2)$.

Classically, a normal-gamma prior is used for parameters μ and λ:

$$\pi_\lambda(\ell|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\ell^{\alpha-1}\exp(-\beta\ell), \qquad \text{for } \ell > 0;$$

$$\pi_{\mu|\lambda}(u|\nu,\lambda,\eta) = (2\,\pi)^{-1/2}\lambda^{1/2}\exp\left(-\frac{\eta\lambda}{2}(t-\nu)^2\right), \qquad \text{for } t\in\mathbb{R}.$$

The parameter η is called the *shrinkage* parameter of the normal prior.

9 Compute the pdf of the normal-gamma prior, i.e. the joint prior pdf of (μ,λ). Display the contour plot of the normal-gamma prior for various values of α, β, ν and η.

To compute the pdf, we simply have to multiply the pdfs defined in both equations above:

$$\begin{aligned} \pi_{\mu,\lambda}(u,\ell|\nu,\eta,\alpha,\beta) &= \pi_{\mu|\lambda}(u|\nu,\lambda,\eta)\cdot\pi_\lambda(\ell|\alpha,\beta),\\ &= (2\,\pi)^{-1/2}\ell^{1/2}\exp\left(-\frac{\eta\ell}{2}(t-\nu)^2\right)\frac{\beta^\alpha}{\Gamma(\alpha)}\ell^{\alpha-1}\exp\left(-\beta\ell\right),\\ &= (2\,\pi)^{-1/2}\frac{\beta^\alpha}{\Gamma(\alpha)}\ell^{\alpha-1/2}\exp\left(-\ell\left(\frac{\eta}{2}(t-\nu)^2+\beta\right)\right). \end{aligned}$$

The contour plot of the normal-gamma prior density can be obtained by

```
alph, beta = (1, 1)
eta, nu = (2, 0)

t1 = np.arange(start=-2.95, stop=3, step=0.05)
t2 = np.arange(start=0.05, stop=4, step=0.05)

prior = [[spst.norm.pdf(i,loc=nu,scale=np.sqrt(1/(eta*j)))*
          spst.gamma.pdf(j,a=alph,scale=1/beta) for i in t1] for j in t2]

t1, t2 = np.meshgrid(t1, t2)
fig2, ax2 = plt.subplots()
CS = ax2.contour(t1, t2, np.asarray(prior), levels=50)

ax2.clabel(CS, inline=1, fontsize=10)
ax2.set_title('$\pi_{\mu,\lambda}(u,\ell)$')
ax2.set_xlabel('$u$')
ax2.set_ylabel('$\ell$')
```

10 Assume that we have observed an iid sample $x_1,\ldots,x_n$ of realizations of a random variable $X \sim \mathcal{N}(\mu,\lambda^{-1})$. Recall the expression for the likelihood function $L(\mu,\lambda)$. Display the contour plot of the likelihood function for a given sample $x_1,\ldots,x_n$.

We have

$$L(\mu, \lambda | x_1, \ldots, x_n) = \prod_{i=1}^{n} p_x(x_i; \mu, \lambda) = (2\pi)^{-n/2} \lambda^{n/2} \exp\left(-\frac{\lambda}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right).$$

11 Show that the posterior distribution for (μ, σ^2) given the sample, λ, α and β is the product of a gamma distribution and a normal distribution. You may drop the computation of the denominator (normalization constant). Display the prior, likelihood, and posterior contours, for various values of n.

We have

$$p_{\mu,\lambda}(u,\ell|x_1,\dots,x_n;\nu,\eta,\alpha,\beta) \propto \ell^{1/2}\exp\left(-\frac{\ell}{2}(\sum_i (x_i-t)^2+\eta(t-\nu)^2)\right)\ell^{\alpha+n/2-1}\exp(-\beta\ell).$$

We can remark that

$$\sum_{i=1}^n (x_i-t)^2 = \sum_{i=1}^n (x_i-\overline{x})^2 + n(\overline{x}-t)^2;$$

$$n(\overline{x}-t)^2+\eta(t-\nu)^2 = (n+\eta)(t-\tilde{\nu})^2+\frac{n\eta(\overline{x}-\nu)^2}{n+\eta}, \quad \text{with } \tilde{\nu}=\frac{n\overline{x}+\eta\nu}{n+\eta};$$

$$\text{thus,} \quad \sum_i (x_i-t)^2+\eta(t-\nu)^2 = \sum_{i=1}^n (x_i-\overline{x})^2+(n+\eta)(t-\tilde{\nu})^2+\frac{n\eta(\overline{x}-\nu)^2}{n+\eta}.$$

This finally gives

$$p_{\mu,\lambda}(u,\ell|x_1,\dots,x_n;\lambda,\alpha,\beta) \propto \mathcal{N}\left(u|\tilde{\nu},((n+\eta)\ell)^{-1}\right) \times \text{gamma}\left(\ell|\alpha+\frac{n}{2},\beta+\frac{1}{2}\sum_{i=1}^n (x_i-\overline{x})^2+\frac{n\eta(\overline{x}-\nu)^2}{2(n+\eta)}\right).$$

The normal-gamma prior is thus conjugate for the normal sampling distribution.
The contour plots can be displayed using the following code.

```
alph, beta = (1, 1)
eta, nu = (2, 0)
mu, sig2, n = (1, 2, 2)
x = spst.norm.rvs(size=n, loc=mu, scale=np.sqrt(sig2))

t1 = np.arange(start=-4.95, stop=5, step=0.1)
t2 = np.arange(start=0.05, stop=4, step=0.05)

prior = [[spst.norm.pdf(i,loc=nu,scale=np.sqrt(1/(eta*j)))*
          spst.gamma.pdf(j,a=alph,scale=1/beta) for i in t1] for j in t2]

ll = [[loglike((i,np.sqrt(1/j)), spst.norm, x) for i in t1] for j in t2]

nutilde = (x.sum()+eta*nu)/(n+eta)
alphtilde = alph+n/2
betatilde = beta+n*x.var()/2+(n*eta*(x.mean()-nu)**2)/(2*(n+eta))
post = [[spst.norm.pdf(i,loc=nutilde,scale=np.sqrt(1/((n+eta)*j)))*
         spst.gamma.pdf(j,a=alphtilde,scale=1/(betatilde))
         for i in t1] for j in t2]

fig, ax = plt.subplots(3, 1, sharex=True, sharey=True, tight_layout=True,
                       figsize=(5,15))
t1_, t2_ = np.meshgrid(t1, t2)

CS = ax[0].contour(t1_, t2_, np.asarray(prior), levels=10)
ax[0].clabel(CS, inline=1, fontsize=10)
ax[0].set_title('$\pi_{\mu,\lambda}(u,\ell)$')
ax[0].set_xlabel('$u$')
ax[0].set_ylabel('$\ell$')

CS = ax[1].contour(t1_, t2_, np.asarray(np.exp(ll)), levels=10)
ax[1].clabel(CS, inline=1, fontsize=10)
ax[1].set_title('$L(\mu=u,\lambda=\ell)$')
ax[1].set_xlabel('$u$')
ax[1].set_ylabel('$\ell$')

CS = ax[2].contour(t1_, t2_, np.asarray(post), levels=10)
ax[2].clabel(CS, inline=1, fontsize=10)
ax[2].set_title('$\pi_{\mu,\lambda}(u,\ell|\cdot)$')
ax[2].set_xlabel('$\mu$')
```