For $a, b > 0$, the distribution $\mathrm{Beta}(a, b)$ is the probability distribution on $[0, 1]$ with density

$$f_{a,b}(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} \ , \qquad \text{where} \qquad B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\, du \ . \tag{1}$$

We admit that its mean and its variance are

$$\frac{a}{a+b} \qquad \text{and} \qquad \frac{ab}{(a+b)^2(a+b+1)} \ . \tag{2}$$

In exercises 2 and 3, $d \geq 1$, $\lambda > 0$, vectors of $\mathbb{R}^d$ are columns, $\langle \cdot, \cdot \rangle$ is the canonical inner product and $\mathbf{I}$ the identity matrix of size d.

Exercise 1 (Beta posteriors and Thompson sampling)**.** Let θ be a random variable with distribution $\mathrm{Beta}(a, b)$ (the prior) and, conditionally on θ, let $Y_1, \ldots, Y_n$ be i.i.d. with distribution $\mathcal{B}(\theta)$. For $y_1, \ldots, y_n \in \{0, 1\}$, we write $s = \sum_{i=1}^n y_i$ and $f = n - s$.

1. Show that $\mathbb{P}(Y_1 = y_1, \ldots, Y_n = y_n \mid \theta) = \theta^s(1-\theta)^f$, and deduce that the distribution of θ conditionally on $\{Y_1 = y_1, \ldots, Y_n = y_n\}$ (the posterior) is $\mathrm{Beta}(a + s, b + f)$.
   Hint: by Bayes' rule, the posterior density is proportional to the product of the prior density and of the conditional probability of the observations.
2. Assume $a = b = 1$ (uniform prior). Compute the mean and the variance of the posterior for $(s, f) = (3, 1)$, then for $(s, f) = (30, 10)$.
3. Assume $a = b = 1$ and $n \geq 1$. Show that the posterior mean is

   $$\frac{1+s}{2+n} = w \cdot \frac{1}{2} + (1-w) \cdot \frac{s}{n} \qquad \text{with} \qquad w = \frac{2}{n+2} \ ,$$

   and that the posterior variance is at most $\frac{1}{4(n+3)}$.
4. Consider two Bernoulli arms with independent uniform priors on their means. Arm 1 has been pulled once, with reward 1, and arm 2 once, with reward 0. Thompson sampling draws $\tilde{\theta}_1$ and $\tilde{\theta}_2$ independently from the two posteriors and plays arm 1 if $\tilde{\theta}_1 > \tilde{\theta}_2$, and arm 2 otherwise. Show that $\mathbb{P}(\tilde{\theta}_2 \leq x) = 2x - x^2$ for any $x \in [0, 1]$, then compute the probability that Thompson sampling plays arm 2.

Exercise 2 (Ridge regression and its recursive update)**.** Let $x_1, \ldots, x_n \in \mathbb{R}^d$ and $r_1, \ldots, r_n \in \mathbb{R}$. We set

$$L(\theta) = \sum_{s=1}^n \big(r_s - \langle \theta, x_s \rangle\big)^2 + \lambda \|\theta\|^2 \ , \qquad \mathbf{V} = \lambda \mathbf{I} + \sum_{s=1}^n x_s x_s^\top \ , \qquad b = \sum_{s=1}^n r_s x_s \ . \tag{3}$$

1. Show that $u^\top \mathbf{V} u = \lambda \|u\|^2 + \sum_{s=1}^n \langle u, x_s \rangle^2$ for any $u \in \mathbb{R}^d$, and deduce that $\mathbf{V}$ is invertible.
2. Show that $\nabla L(\theta) = 2(\mathbf{V}\theta - b)$, and deduce that L has a unique minimizer, $\hat{\theta} = \mathbf{V}^{-1} b$.
3. Let $x \in \mathbb{R}^d$ and $q = x^\top \mathbf{V}^{-1} x$. Show that

   $$\big(\mathbf{V} + x x^\top\big)^{-1} = \mathbf{V}^{-1} - \frac{\mathbf{V}^{-1} x x^\top \mathbf{V}^{-1}}{1+q} \ . \tag{4}$$

   Hint: multiply the right-hand side by $\mathbf{V} + x x^\top$ on the left.

4. Deduce that $x^\top \left(\mathbf{V} + xx^\top\right)^{-1} x = \dfrac{q}{1+q}$.
5. Let $x \neq 0$, $\mathbf{V}_0 = \lambda \mathbf{I}$ and $\mathbf{V}_k = \mathbf{V}_{k-1} + xx^\top$ for $k \geq 1$ (the same vector is observed at each step). Show that, for any $k \geq 0$,

$$x^\top \mathbf{V}_k^{-1} x = \frac{\|x\|^2}{\lambda + k\|x\|^2} \ .$$

Exercise 3 (LinUCB on an example)**.** With the notation (3) and $\hat{\theta} = \mathbf{V}^{-1} b$, the index of LinUCB with parameter $\beta > 0$ for an arm with feature vector $x \in \mathbb{R}^d$ is

$$U(x) = \langle \hat{\theta}, x \rangle + \beta \sqrt{x^\top \mathbf{V}^{-1} x} \ . \tag{5}$$

1. Assume $d = K$ and that the feature vector of arm a is e_a, the a-th vector of the canonical basis. Arm a has been pulled N_a times, and the sum of its rewards is S_a. Show that $\mathbf{V}$ is the diagonal matrix with entries $\lambda + N_a$, and that

$$U(e_a) = \frac{S_a}{\lambda + N_a} + \frac{\beta}{\sqrt{\lambda + N_a}} \ .$$

In the following questions, $d = 2$, $\lambda = 1$, $\beta = 1$, and there are three arms with feature vectors

$$x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ , \qquad x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ , \qquad x_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \ .$$

Arm 1 has been pulled four times, with rewards $1, 1, 0, 0$, and arm 2 once, with reward 0. Arm 3 has not been pulled. The following approximate values may be used: $\sqrt{0.2} \approx 0.447$, $\sqrt{0.5} \approx 0.707$, $\sqrt{0.7} \approx 0.837$, $\sqrt{0.176} \approx 0.420$ and $\sqrt{0.412} \approx 0.642$.

2. Compute $\mathbf{V}$, b and $\hat{\theta}$.
3. Compute $U(x_1)$, $U(x_2)$ and $U(x_3)$, and deduce the arm played by LinUCB.
4. Assume that arm 3 is played, and let $\mathbf{V}' = \mathbf{V} + x_3 x_3^\top$. Using (4), compute $x_1^\top \mathbf{V}'^{-1} x_1$ and $x_3^\top \mathbf{V}'^{-1} x_3$, and the new values of the bonuses of arms 1 and 3.