Elements of Bayesian statistics

nicolas.jouvin@inrae.fr

Reminder on some useful distributions

- **The Dirichlet distribution** is for discrete probability vector $(\theta_1, \ldots, \theta_K)$. It is a continuous distribution over the simplex $\mathcal{S}_K := \left\{\theta \in \mathbb{R}^K : \theta_k \geq 0, \sum_k \theta_k = 1\right\}$ with p.d.f.

$$Dir(\theta \mid \alpha_1, \ldots, \alpha_K) = \frac{1}{B(\alpha)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}, \quad \text{with } B(\alpha) := \frac{\prod_k \Gamma(\alpha_k)}{\Gamma(\sum_k \alpha_k)} \text{ the normalizing constant.}$$

  $\alpha = (\alpha_1, \ldots, \alpha_K)$ are hyper-parameters, more info on Wikipedia. When $K = 2$, it is equivalent to the Beta distribution.

- **The multinomial distribution** over $\mathbb{N}^K$ basically generalizes Binomial distribution when there are $K \geq 2$ possible outcomes. It can be thought of as the probability of counts for each side of a K-sided dice rolled L times. It a multivariate *discrete* probability distribution on $\mathcal{C} = \{(x_1, \ldots, x_K) \in \mathbb{N}^L : \sum_k x_k = L\}$ with mass function

$$\mathcal{M}_K(x_1, \ldots, x_K \mid \theta, L) = \mathbb{P}(X = (x_1, \ldots, x_K) \mid \theta, L) = \mathbf{1}_{(x_1, \ldots, x_K) \in \mathcal{C}} \frac{L!}{\prod_{k=1}^{K} x_k!} \prod_{k=1}^{K} \theta_k^{x_k}$$

  Here, $\theta_1, \ldots, \theta_K$ are the probability for each side of the dice.

  N.B. when $L = 1$ (one draw), we call the multinomial a *categorical* distribution.

- **The gamma distribution** over $\mathbb{R}_+$ is driven by two hyper-parameters $a, b > 0$ called *shape* and *rate*. Its p.d.f. is written as

$$\mathcal{G}(\theta \mid a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \propto \theta^{a-1} e^{-b\theta}.$$

  Special cases are the exponential $(a = 2)$ & $\chi^2(k)$ $(a = k/2$ and $b = 1/2)$ distributions.

- **The Poisson distribution** is a discrete probability measure over $\mathbb{N}$ with intensity parameter $\theta > 0$ and p.d.f.

$$\forall x \in \mathbb{N}, \quad \mathbb{P}(X = x \mid \theta) = \mathcal{P}(x \mid \theta) = e^{-\theta} \frac{\theta^x}{x!}.$$

Exercise 1 *(Conjugate posteriors).* Derive the posterior distribution of $\theta \mid X = x$ for the following models

1. *Dirichlet-Multinomial:* $\theta \sim Dir_K(\alpha_1, \ldots, \alpha_K)$ and $X_1, \ldots, X_n \mid \theta \overset{i.i.d.}{\sim} \mathcal{M}_K(1, \theta)$.
2. *Gamma-Poisson*: $\theta \sim Gamma(a, b)$ and $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{P}(\theta)$

Exercise 2 *(Bayesian decision theory).* Derive the form of the Bayes estimator

$$\hat{\theta} \in \arg\min_{\eta} \left\{ \mathbb{E}_{\theta \sim \pi(\cdot \mid x)} \left[C(\eta, \theta) \right] = \int_{\Theta} C(\eta, \theta) \pi(\theta \mid x) \, \mathrm{d}\theta \right\},$$

for the following cost functions

1. L^2 loss: $C(\eta,\theta) = (\eta-\theta)^2$, $\Theta \subset \mathbb{R}^d$
2. L^1 loss: $C(\eta,\theta) = |\eta-\theta|$ when $d = 1$
3. $0-1$ loss: $C(\eta,\theta) = \mathbf{1}_{\eta\neq\theta}$, when Θ is finite.
4. Linear-Exponential loss: for $a > 0$, $C(\eta,\theta) = e^{a(\eta-\theta)} - a(\eta-\theta) - 1$, $\Theta \subset \mathbb{R}$

Exercise 3 *(Beta-Binomial model).* We recall the Beta-Binomial model (seen in the slides) $\theta \sim Beta(a,b)$ and $X_1,\ldots,X_n \overset{i.i.d.}{\sim} Ber(\theta)$, with posterior $\theta \mid X \sim Beta(a+\sum_i X_i, b+n-\sum_i X_i)$.
Show that

1. the MLE is $\theta^{MLE} = \sum_i X_i/n$.
2. the posterior mean can be written as a convex combination of the MLE and the prior expectation
$$\mathbb{E}[\theta \mid X] = \lambda_n \cdot \theta^{MLE} + (1-\lambda_n)\cdot \mathbb{E}_{prior}[\theta]$$
3. What do you deduce of the prior's impact on the posterior mean as $n \to +\infty$?