BML: exercise sheet

Stars indicate the difficulty level, from 1 to 3. One star means that everyone should be able to do it without too much effort.

1 Lecture: Bayesian nonparametrics

1.1 Combinatorial properties of K_n for Dirichlet process ($\star$)

Let K_n be the number of clusters observed when drawing n observations from a Dirichlet process with concentration parameter $\alpha \in \mathbb{R}_+$.

1. Show that

$$\mathbb{E}[K_n] = \sum_{i=0}^{n-1} \frac{\alpha}{\alpha + i} \quad \text{and} \quad \mathrm{Var}(K_n) = \sum_{i=0}^{n-1} \frac{\alpha i}{(\alpha + i)^2}.$$

2. Show the following large n asymptotics for the expectation and variance of K_n:

$$\mathbb{E}[K_n] \sim \alpha \log n \quad \text{and} \quad \mathrm{Var}(K_n) \sim \alpha \log n.$$

1.2 Combinatorial properties of K_n for Pitman–Yor process ($\star\star$)

Let K_n be the number of clusters observed when drawing n observations from a Pitman–Yor process with discount parameter $\sigma \in (0, 1)$ and concentration parameter $\alpha \in \mathbb{R}_+$.

1. Show that

$$\mathbb{E}[K_{n+1}] = \frac{\alpha}{n + \alpha} + \frac{\sigma + \alpha + n}{n + \alpha} \mathbb{E}\left[K_n\right].$$

   Hint: use the PY predictive distribution and a conditional expectation to get this iterative formula from n to $n + 1$.

2. Deduce that

$$\mathbb{E}[K_n] = \sum_{i=0}^{n-1} \frac{(\alpha + \sigma)_i}{(\alpha + 1)_i},$$

   where $(x)_n = x(x + 1) \ldots (x + n - 1)$.

3. Show the following large n asymptotics for the expectation of K_n:

$$\mathbb{E}[K_n] \sim \frac{\Gamma(\alpha+1)}{\sigma\Gamma(\alpha+\sigma)} n^{\sigma}.$$

4. Show that the following recursive formula holds for the variance of K_n:

$$\operatorname{Var}(K_{n+1}) = \operatorname{Var}(K_n)\frac{n+\alpha+2\sigma}{n+\alpha} + \frac{(\sigma\mathbb{E}[K_n]+\alpha)(n-\sigma\mathbb{E}[K_n])}{(n+\alpha)^2}. \tag{1}$$

   Hint: use the law of total variance.

5. Derive again the simpler expression of expectation and variance of K_n in the Dirichlet process case (by setting $\sigma = 0$ in the above formulas).