

Homework 1

Statistical Methods in Applied Computer Science

DD2447

Jim Holmström 890503-7571

jimho@kth.se

November 5, 2012

Exercise 2.16 Mean, mode, variance for the beta distribution
Suppose $\theta \sim \text{Beta}(a, b)$. Derive the mean, mode and variance.

Solution. Denote $\text{Beta}(a, b) = \text{Beta}_{a,b}$ to simplify the expressions later on. The mean of a random variable sampled from $\text{Beta}(a, b)$ is defined as:

$$\mathbb{E} \theta = \int_{\Omega} \theta' \text{Beta}_{a,b}(\theta') d\theta' \quad (1)$$

where

$$\text{Beta}_{a,b}(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}, B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (2)$$

and Ω is the support $[0, 1]$

Lemma 0.1.

$$B(a+1, b) = B(b, a) \bigwedge B(a+1, b) = \frac{a}{a+b} B(a, b) \quad (3)$$

Proof. The first relation is given by the symmetry in the definition of B together with a commutative multiplication. For the second relation we first need to show the well known identity

$$\Gamma(z+1) = z\Gamma(z) \quad (4)$$

By the definition of $\Gamma(z) \triangleq \int_0^{\infty} u^{z-1} e^{-u} du$ and by using partial integration we get

$$\Gamma(z+1) = \int u^z e^{-u} du = \left[-u^z e^{-u} \right]_{u=0}^{\infty} + \int_0^{\infty} z u^{z-1} e^{-u} du = z \int_0^{\infty} u^{z-1} e^{-u} du = z\Gamma(z) \quad (5)$$

Having this relation we can show that

$$B(a+1, b) = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} = \frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)} = \frac{a}{a+b}B(a, b) \quad (6)$$

□

Starting out by deriving the mean:

$$\begin{aligned} E\theta &= \int \theta' Beta_{a,b}(\theta') d\theta' = \int \frac{B(a+1, b) Beta_{a+1,b}(\theta')}{B(a, b)} d\theta' = \\ &= \frac{B(a+1, b)}{B(a, b)} \int Beta_{a+1,b}(\theta') d\theta' = \frac{B(a+1, b)}{B(a, b)} = \{\text{Using Lemma 0.1}\} = \\ &= \frac{aB(a, b)}{(a+b)B(a, b)} = \frac{a}{a+b} \end{aligned} \quad (7)$$

Next we derive the mode which is the most occurring θ

$$Mode(\theta) = \underset{\theta'}{argmax}(Beta_{a,b}(\theta')) \quad (8)$$

To find it we use

$$\frac{\partial Beta_{a,b}(\theta)}{\partial \theta} = 0, \left. \frac{\partial^2 Beta_{a,b}(\theta)}{\partial \theta^2} \right|_{\theta: \frac{\partial Beta_{a,b}(\theta)}{\partial \theta} = 0} < 0 \quad (9)$$

$$\begin{aligned} 0 &= B(a, b) \frac{\partial Beta_{a,b}(\theta)}{\partial \theta} = \{\text{prod. rule}\} = \\ &= (a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2} \end{aligned} \quad (10)$$

which gives us the equation

$$(a-1)\theta^{a-2}(1-\theta)^{b-1} = (b-1)\theta^{a-1}(1-\theta)^{b-2} \quad (11)$$

$$(a-1)(1-\theta) = (b-1)\theta \quad (12)$$

$$\theta = \frac{1}{1 + \frac{b-1}{a-1}} = \frac{a-1}{a-1+b-1} = \frac{a-1}{a+b-2} \quad (13)$$

$$\begin{aligned} \frac{\partial^2 Beta_{a,b}(\theta)}{\partial \theta^2} &= \frac{\partial((10))}{\partial \theta} = \frac{\partial(a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2}}{\partial \theta} = \\ &= \{\text{prod. rule}\} = (a-1)(a-2)\theta^{a-3}(1-\theta)^{b-1} - (a-1)(b-1)\theta^{a-2}(1-\theta)^{b-2} \\ &\quad - (b-1)(a-1)\theta^{a-2}(1-\theta)^{b-2} + (b-1)(b-2)\theta^{a-1}(1-\theta)^{b-3} \end{aligned} \quad (14)$$

$$\left. \frac{\partial^2 \text{Beta}_{a,b}(\theta)}{\partial \theta^2} \right|_{\theta = \frac{a-1}{a+b-2}} = \dots = \frac{\left(\frac{a-1}{a+b-2}\right)^a \left(\frac{b-1}{a+b-2}\right)^b (a+b-2)^5}{(a-1)^2(b-1)^2} < 0 \quad (15)$$

$$\frac{\left(\frac{a-1}{a+b-2}\right)^a \left(\frac{b-1}{a+b-2}\right)^b (a+b-2)^5}{(a-1)^2(b-1)^2} < 0 \quad (16)$$

$$\left(\frac{a-1}{a+b-2}\right)^a \left(\frac{b-1}{a+b-2}\right)^b (a+b-2)^5 < 0 \quad (17)$$

$$(a-1)^a(b-1)^b(a+b-2)^{\frac{5}{a+b}} < 0 \quad (18)$$

we can from here see that $a, b > 1$, in the other cases the distribution will not have a well defined mode.

$$\therefore \text{Mode}(\theta) = \frac{a-1}{a+b-2} \bigwedge a, b > 1 \quad (19)$$

Finally we deal with the variance:

$$\begin{aligned} \text{Var}(\theta) &= \text{E} \theta^2 - (\text{E} \theta)^2 = \int \theta'^2 \text{Beta}_{a,b}(\theta') d\theta' - (\text{E} \theta)^2 = \\ &= \frac{B(a+2, b)}{B(a, b)} - (\text{E} \theta)^2 = \{\text{Using Lemma 0.1}\} = \frac{(a+1)}{(a+b+1)} \frac{a}{(a+b)} - \left(\frac{a}{a+b}\right)^2 = \\ &= \frac{a}{a+b} \left(\frac{(a+1)(a+b) - a(a+b+1)}{(a+b)(a+b+1)} \right) = \frac{ab}{(a+b)^2(a+b+1)} \end{aligned} \quad (20)$$

Exercise 3.6 MLE for the Poisson distribution

The Poisson pmf is defined as $\text{Poi}(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$ for $x \in \{0, 1, 2, \dots\}$ where $\lambda > 0$ is the rate parameter. Derive the MLE.

Solution. The data $D = \{x_i\}$ is *i.i.d* $\bigwedge x_i \sim \text{Poi}(\lambda)$

$$p(x_i|\lambda) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \quad (21)$$

The likelihood of D :

$$L(\lambda) = p(\{x_i\}|\lambda) = \prod_i p(x_i|\lambda) = \prod_i e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \quad (22)$$

The log-likelihood:

$$\ell(\lambda) = \log\left(\prod_i e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}\right) = \sum_i (-\lambda + x_i \log \lambda - \log x_i) \quad (23)$$

Find the maximum likely λ by finding the maximum log-likely λ which will coincide since log is a strictly increasing function.

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = 0, \frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} < 0 \quad (24)$$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \sum_i \left(-1 + \frac{x_i}{\lambda} \right) = -|\{x_i\}| + \frac{\sum x_i}{\lambda} = 0 \quad (25)$$

$$\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} = -\frac{\sum x_i}{\lambda^2} < 0 \quad (26)$$

Remark. If $x_i = 0$ we will have problems with the MLE, but for sufficient number of data points this event will have probability zero compared to having data with $\exists i : x_i > 0$ and can thus be ignored in the analysis.

Which gives the MLE for λ :

$$\lambda = \frac{\sum x_i}{|\{x_i\}|} (= \mathbb{E} x) \quad (27)$$

Exercise 3.7 Bayesian analysis of the Poisson distribution

In exercise 3.6, we defined the Poisson distribution with rate λ and derived its MLE. Here we perform a conjugate Bayesian analysis.

- Derive the posterior $p(\lambda|D)$ assuming a conjugate prior $p(\lambda) = Ga(\lambda|a, b) \propto \lambda^{a-1} e^{-\lambda b}$.
Hint: the posterior is also a Gamma distribution.
- What does the posterior mean tend to as $a \rightarrow 0$ and $b \rightarrow 0$? (Recall that the mean of a $Ga(a, b)$ distribution is a/b .)

Solution. $Ga(a, b)$ is denoted $\Gamma(a, b)$ and $a, b > 0$ of a Γ -distribution.

$$\begin{aligned} p(\lambda|D) &\propto p(D|\lambda)p(\lambda) = \lambda^{a-1} e^{-\lambda b} \prod_i \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \propto \lambda^{a-1} e^{-\lambda b} e^{-|D|\lambda} \prod_i \lambda^{x_i} = \\ &= e^{-|D|\lambda} e^{-\lambda} \lambda^{a-1} \lambda^{\sum x_i} = e^{-(|D|+b)\lambda} \lambda^{\sum x_i + a - 1} \propto \Gamma\left(\sum x_i + a, |D| + b\right) \end{aligned} \quad (28)$$

with the parameters for the Γ -function being $a' = \sum x_i + a \geq a > 0$ and $b' = |D| + b \geq b > 0$ together with the fact that $p(\lambda|D)$ is a distribution, that is properly normalized, it will result in not only being proportional but equally distributed:

$$p(\lambda|D) = \Gamma\left(\sum x_i + a, |D| + b\right) \quad (29)$$

still acting as a distribution, the samples

$$X \sim \Gamma\left(\sum x_i + a, |D| + b\right) \quad (30)$$

in the limit the E becomes

$$\lim_{a,b \rightarrow 0} E X = \lim_{a,b \rightarrow 0} \frac{\sum x_i + a}{|D| + b} = \frac{\sum x_i}{|D|} (= E x) \quad (31)$$