# 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 $\theta \sim Beta(a,b)$. Derive the mean, mode and variance.

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

$$\mathrm{E}\,\theta = \int_{\Omega} \theta' Beta_{a,b}(\theta') d\theta' \tag{1}$$

where

$$Beta_{a,b}(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}, B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \tag{2}$$

and $\Omega$ is the support $[0,1]$

*Lemma* 0.1.

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

*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

$$\Gamma(z+1) = z\Gamma(z) \tag{4}$$

By the definition of $\Gamma(z) \triangleq \int\limits_0^\infty u^{z-1}e^{-u}du$ and by using partial integration we get

$$\Gamma(z+1) = \int u^z e^{-u} du = \big[-u^z e^{-u}\big]_{u=0}^{\infty} + \int\limits_0^\infty z u^{z-1} e^{-u} du = z \int\limits_0^\infty u^{z-1} e^{-u} du = z\Gamma(z) \tag{5}$$

Having this relation we can show that

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

□

Starting out by deriving the mean:

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

Next we derive the mode which is the most occurring $\theta$

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

To find it we use

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

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

which gives us the equation

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

$$(a-1)(1-\theta) = (b-1)\theta \tag{12}$$

$$\theta = \frac{1}{1+\frac{b-1}{a-1}} = \frac{a-1}{a-1+b-1} = \frac{a-1}{a+b-2} \tag{13}$$

$$\begin{aligned} \frac{\partial^2 Beta_{a,b}(\theta)}{\partial\theta^2} = \frac{\partial((10))}{\partial\theta} = \frac{\partial(a-1)\theta^{a-2}(1-\theta)^{b-1} - (b-1)\theta^{a-1}(1-\theta)^{b-2}}{\partial\theta} = \\ = (a-1)(a-2)\theta^{a-3}(1-\theta)^{b-1} - (a-1)(b-1)\theta^{a-2}(1-\theta)^{b-2} \\ -(b-1)(a-1)\theta^{a-2}(1-\theta)^{b-2} + (b-1)(b-2)\theta^{a-1}(1-\theta)^{b-3} \end{aligned} \tag{14}$$

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

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

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

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

we can easily see from the last power-term that $a+b>2$. This makes $(a+b-2)^c$ will be non-negative $\forall c$ and can thus be cancelled out.

$$(a-1)^a(b-1)^b < 0 \tag{19}$$

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

$$\therefore Mode(\theta) = \frac{a-1}{a+b-2}\bigwedge a,b>1 \tag{20}$$

Finally we deal with the variance:

$$\begin{aligned}
\mathrm{Var}\,(\theta) = \mathrm{E}\,\theta^2 - (\mathrm{E}\,\theta)^2 = \int \theta'^2 Beta_{a,b}(\theta')d\theta' - (\mathrm{E}\,\theta)^2 = \\
= \frac{B(a+2,b)}{B(a,b)} - (\mathrm{E}\,\theta)^2 = \{\text{Using Lemma 0.1}\} = \frac{(a+1)}{(a+b+1)}\frac{a}{(a+b)} - \left(\frac{a}{a+b}\right)^2 = \\
= \frac{a}{a+b}\left(\frac{(a+1)(a+b)-a(a+b+1)}{(a+b)(a+b+1)}\right) = \frac{ab}{(a+b)^2(a+b+1)}
\end{aligned} \tag{21}$$

**Exercise 3.6** MLE for the Poisson distribution
The Poisson pmf is defined as $Poi(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$ for $x \in \{0,1,2,..\}$ where $\lambda > 0$ is the rate parameter. Derive the MLE.

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

$$p(\,x_i|\lambda\,) = e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \tag{22}$$

The likelihood of $D$:

$$L(\lambda) = p(\,\{x_i\}|\lambda\,) = \prod_i p(\,x_i|\lambda\,) = \prod_i e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \tag{23}$$

The log-likelihood:

$$\ell(\lambda) = \log(\prod_i e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}) = \sum_i (-\lambda + x_i \log \lambda - \log x_i) \tag{24}$$

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

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = 0, \frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} < 0 \tag{25}$$

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \sum_i \left( -1 + \frac{x_i}{\lambda} \right) = -|\{x_i\}| + \frac{\sum x_i}{\lambda} = 0 \tag{26}$$

$$\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} = -\frac{\sum x_i}{\lambda^2} < 0 \tag{27}$$

*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 $\lambda$:

$$\lambda = \frac{\sum x_i}{|\{x_i\}|} (= \mathrm{E}\, x) \tag{28}$$

**Exercise 3.7** Bayesian analysis of the Poisson distribution
In exercise 3.6, we defined the Poisson distribution with rate $\lambda$ and derived its MLE. Here we perform a conjugate Bayesian analysis.

**a.** 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.

**b.** What does the posterior mean tend to as $a \to 0$ and $b \to 0$? (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 $\Gamma$-distribution.

$$\begin{aligned} p(\lambda|D) \propto p(D|\lambda)p(\lambda) &= \lambda^{a-1} e^{-\lambda b} \prod_i (e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}) \propto \lambda^{a-1} e^{-\lambda b} e^{-|D|\lambda} \prod_i \lambda^{x_i} = \\ &= e^{-|D|\lambda} e^{-\lambda} \lambda^{a-1} \lambda^{\sum x_i} = e^{-(|D|+b)\lambda} \lambda^{\sum x_i + a - 1} \propto \Gamma\Big( \sum x_i + a, |D| + b \Big) \end{aligned} \tag{29}$$

with the parameters for the $\Gamma$-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:

$$p(\lambda|D) = \Gamma\Big( \sum x_i + a, |D| + b \Big) \tag{30}$$

still acting as a distribution, the samples

$$X \sim \Gamma\left( \sum x_i + a, |D| + b \right) \tag{31}$$

in the limit the E becomes

$$\lim_{a,b \to 0} \mathrm{E}\, X = \lim_{a,b \to 0} \frac{\sum x_i + a}{|D| + b} = \frac{\sum x_i}{|D|} (= Ex) \tag{32}$$