British Journal of Mathematics & Computer Science
16(3): 1-21, 2016, Article no.BJMCS.25731
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Beta Likelihood Estimation and Its Application to Specify Prior Probabilities in Bayesian Network

Loc Nguyen[1*]

[1]*Sunflower Soft Company, Ho Chi Minh City, Vietnam.*

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/BJMCS/2016/25731
Editor(s):
(1) H. M. Srivastava, Department of Mathematics and Statistics, University of Victoria, Canada.
(2) Paul Bracken, Department of Mathematics, The University of Texas-Pan American Edinburg, TX 78539, USA.
Reviewers:
(1) Radosław Jedynak, Kazimierz Pulaski University of Technology and Humanities, Poland.
(2) Anonymous, USA.
(3) S. Zimeras, University of the Aegean, Greece.
Complete Peer review History: http://sciencedomain.org/review-history/14364

Received: 17th March 2016
Accepted: 20th April 2016
Published: 27th April 2016

Method Article

Abstract

Maximum likelihood estimation (MLE) is a popular technique of statistical parameter estimation. When random variable conforms beta distribution, the research focuses on applying MLE into beta density function. This method is called beta likelihood estimation, which results out useful estimation equations. It is easy to calculate statistical estimates based on these equations in case that parameters of beta distribution are positive integer numbers. Essentially, the method takes advantages of interesting features of functions gamma, digamma, and trigamma. An application of beta likelihood estimation is to specify prior probabilities in Bayesian network.

Keywords: Maximum likelihood estimation; beta distribution; beta likelihood estimation; gamma function.

1 Introduction

1.1 Introduction to maximum likelihood estimation

Let Θ and X be the hypothesis and observation variable, respectively. Suppose $x_1, x_2, \ldots, x_n$ are instances of variable X in training data and they are observed independently. According multiplication rule in probability

**Corresponding author: E-mail: ng_phloc@yahoo.com;*

theory, the likelihood function $L(\Theta)$ is the joint probability which is the product of condition probabilities of instances x_i, given hypothesis variable Θ (Lynch, 2007, p. 36). Equation 1 expresses the likelihood function $L(\Theta)$ with regard to variable Θ.

$$L(\Theta) = P(X \mid \Theta) = \prod_{i=1}^{n} P(x_i \mid \Theta) \tag{1}$$

Where $P(x_i|\Theta)$ is the conditional probability of instance x_i given the hypothesis Θ. Suppose $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$ is the vector of parameters specifying arbitrary distribution of X, it is required to estimate the parameter vector and its standard deviation so that the likelihood function takes maximum value. Thus, this method is called maximum likelihood estimation (MLE). The parameter vector that maximizes likelihood function is called *parameter vector estimate* denoted $\widehat{\Theta}$, as shown in equation 2.

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmax}}\, L(\Theta) = \underset{\Theta}{\operatorname{argmax}} \left(\prod_{i=1}^{n} P(x_i \mid \Theta) \right) \tag{2}$$

The natural logarithm of $L(\Theta)$ is called log-likelihood denoted $LnL(\Theta)$, as shown in equation 3 [1, p. 38].

$$\begin{aligned} LnL(\Theta) &= ln\left(\prod_{i=1}^{n} P(x_i \mid \Theta) \right) = \sum_{i=1}^{n} ln\big(P(x_i \mid \Theta)\big) \\ \widehat{\Theta} &= \underset{\Theta}{\operatorname{argmax}}\, LnL(\Theta) = \underset{\Theta}{\operatorname{argmax}} \left(\sum_{i=1}^{n} ln\big(P(x_i \mid \Theta)\big) \right) \end{aligned} \tag{3}$$

Where *ln*(.) denotes natural logarithm function.

The essence of maximizing the likelihood function is to find the peak of the curve of $LnL(\Theta)$ [1, p. 38]. This can be done by setting the first-order partial derivative of $LnL(\Theta)$ with respect to each parameter θ_i to 0 and solving this equation to find out parameter θ_i. The number of equations corresponds with the number of parameters. If all parameters are found, in other words, the parameter vector estimate $\widehat{\Theta} = \{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k\}$ is defined then the distribution of X is known clearly. Each $\hat{\theta}_i$ is also called a *parameter estimate*.

The accuracy of parameter estimator is measured by its standard error [2, p. 225] and thus; another important issue is how to determine the standard error when we have already computed all parameters and standard error is standard deviation of parameter estimator. It is very fortunate when the second-order derivative of the log-likelihood function denoted $\frac{\partial^2 LnL}{\partial\Theta\partial\Theta^T}$ can be computed and it is used to determine the variances of parameters. If there is only one parameter, the second-order derivative $\frac{\partial^2 LnL}{\partial\Theta\partial\Theta^T}$ is scalar, otherwise it is a matrix so-called Hessian matrix. The negative expectation of Hessian matrix is called the *information matrix* which in turn is inverted so as to construct *co-variance matrix* denoted $Var(\Theta)$ [1, p. 40]. Equation 4 specifies the co-variance matrix of parameter vector Θ.

$$Var(\Theta) = \left(-E\left(\frac{\partial^2 LnL}{\partial\Theta\partial\Theta^T} \right) \right)^{-1} \tag{4}$$

Elements on co-variance matrix diagonal are variances of the parameters and the square root of each variance is a standard error. Note that $\left(-E\left(\frac{\partial^2 LnL}{\partial\Theta\partial\Theta^T} \right) \right)^{-1}$ is exactly so-called Cramer-Rao lower bound of co-variance matrix but we can consider approximately $\left(-E\left(\frac{\partial^2 LnL}{\partial\Theta\partial\Theta^T}\right)\right)^{-1}$ as co-variance matrix when $\widehat{\Theta}$ is derived from likelihood function and $\widehat{\Theta}$ is unbiased estimator [3, p. 11]. Please read [1, pp. 35-43] and [3] for more details about MLE.

In case that variable X conforms beta distribution, MLE is called *beta likelihood estimation*. Next section focuses on how to apply MLE into beta distribution, which is the main purpose of this research.

2 Beta Likelihood Estimation

Before mentioning how to apply MLE into beta distribution, it is important to research aspects of beta distribution. Suppose variable X conforms beta distribution. Equation 5 specifies the beta density function of X as follows:

$$f(X;a,b) = \beta(X;a,b) = beta(X;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} X^{a-1}(1-X)^{b-1} \tag{5}$$

Where Γ(.) denotes gamma function which is expressed as follows:

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t$$

Note that $e^{(.)}$ and *exp*(.) denote exponent function and $e \approx 2.71828$ is Euler's number.

Fig. 1 [4] shows beta density function with various parameters (a=2, b=2), (a=4, b=2), and (a=2, b=4).

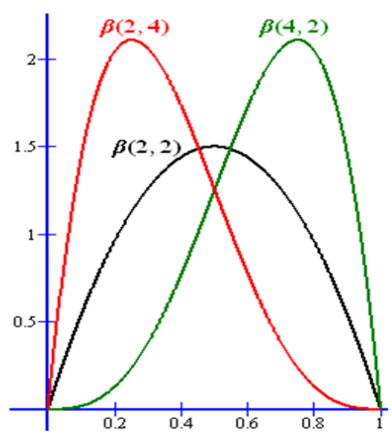

Fig. 1. Beta density functions with various parameters *a* and *b*

Beta density function is based on gamma function and there is another so-called *digamma function* is also defined via gamma function. Equation 6 is definition of digamma function $\psi(x)$. We will know later that beta density function is also relevant to digamma function.

$$\psi(x) = \mathrm{d}\left(ln\left(\Gamma(x)\right)\right) = \frac{\Gamma'(x)}{\Gamma(x)} \tag{6}$$

Note that *ln*(.) denotes natural logarithm function. According to equation 6, digamma function is the derivative of natural logarithm of gamma function.

The integral form of digamma function is specified by equation 7 [5, p. 114]:

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \int_0^{+\infty} \left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1 - e^{-t}} \right) \mathrm{d}t \tag{7}$$

Suppose variable x is non-zero, we have:

$$\psi(x+1) = \int_0^{+\infty} \left(\frac{e^{-t}}{t} - \frac{e^{-(x+1)t}}{1 - e^{-t}} \right) \mathrm{d}t = \int_0^{+\infty} \left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1 - e^{-t}} + e^{-xt} \right) \mathrm{d}t$$
$$= \int_0^{+\infty} \left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1 - e^{-t}} \right) \mathrm{d}t + \int_0^{+\infty} e^{-xt} \mathrm{d}t = \psi(x) - \frac{1}{x} e^{-xt} \Big|_0^{+\infty}$$
$$= \psi(x) + \frac{1}{x}$$

Briefly, the recurrence equation of digamma function is specified by equation 8 [6].

$$\psi(x+1) = \psi(x) + \frac{1}{x} \tag{8}$$

Equation 8 shows recurrence relation [6] of digamma function, which implicates that it is very easy to compute $\psi(x)$ if variable x is positive integer. Thus, it is necessary to calculate $\psi(1)$ which is the evaluation of digamma function at starting positive point 1, we have:

$$\psi(1) = \int_0^{+\infty} \left(\frac{e^{-t}}{t} - \frac{e^{-t}}{1 - e^{-t}} \right) \mathrm{d}t = \int_0^{+\infty} \frac{e^{-t}}{t} dt - \int_0^{+\infty} \frac{e^{-t}}{1 - e^{-t}} \mathrm{d}t$$
$$= \int_0^{+\infty} e^{-t} \mathrm{d}(lnt) - \int_0^{+\infty} \mathrm{d}\big(ln(1 - e^{-t})\big)$$
$$= e^{-t} lnt \Big|_0^{+\infty} + \int_0^{+\infty} e^{-t} lnt \mathrm{d}t - ln(1 - e^{-t}) \Big|_0^{+\infty}$$
$$= \lim_{t \to +\infty} (e^{-t} lnt) - \lim_{t \to 0} (e^{-t} lnt) - \gamma - \lim_{t \to +\infty} ln(1 - e^{-t}) + \lim_{t \to 0} ln(1 - e^{-t})$$

(due to Euler-Mascheroni constant $\gamma = -\int_0^{+\infty} e^{-t} lnt \mathrm{d}t$)

$$= -\gamma + \lim_{t \to +\infty} (e^{-t} lnt) - \lim_{t \to 0} (e^{-t} lnt) + \lim_{t \to 0} ln(1 - e^{-t})$$
$$= -\gamma + \lim_{t \to +\infty} \frac{lnt}{e^t} + \lim_{t \to 0} (ln(1 - e^{-t}) - e^{-t} lnt)$$
$$= -\gamma + \lim_{t \to +\infty} \frac{lnt}{e^t} + \lim_{t \to 0} ln \frac{e^{ln(1 - e^{-t})}}{e^{e^{-t} lnt}}$$

(due to transformation with regard to indeterminate form [7])

$$= -\gamma + \lim_{t \to +\infty} \frac{lnt}{e^t} + \lim_{t \to 0} ln \frac{1 - e^{-t}}{t^{e^{-t}}}$$
$$= -\gamma + \lim_{t \to +\infty} \frac{lnt}{e^t} + ln \left(\lim_{t \to 0} \frac{1 - e^{-t}}{t^{e^{-t}}} \right)$$

$$= -\gamma + \lim_{t\to+\infty}\frac{\mathrm{d}(lnt)}{\mathrm{d}(e^t)} + ln\left(\lim_{t\to 0}\frac{1-e^{-t}}{t^{e^{-t}}}\right)$$

(using L'Hôpital's rule by taking derivatives of both numerator and denominator [7])

$$= -\gamma + \lim_{t\to+\infty}\frac{1}{te^t} + ln\left(\lim_{t\to 0}\frac{1-e^{-t}}{t^{e^{-t}}}\right)$$
$$= -\gamma + ln\left(\lim_{t\to 0}\frac{1-e^{-t}}{t^{e^{-t}}}\right)$$
$$= -\gamma + ln\left(\lim_{t\to 0}\frac{e^{-t}}{g^{'}(t)}\right) \text{ where } g(t) = t^{e^{-t}}$$

Note that $\gamma \approx 0.577215$ is Euler-Mascheroni constant, please read [8] for more detailed about Euler-Mascheroni constant.

$$\gamma = -\int_0^{+\infty} e^{-x}lnx\mathrm{d}x = \lim_{n\to+\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - ln(n)\right) \approx 0.577215$$

We have:

$$ln(g(t)) = e^{-t}lnt \Longrightarrow \frac{g^{'}(t)}{g(t)} = \frac{e^{-t}(1-tlnt)}{t} \Longrightarrow g^{'}(t) = \frac{t^{e^{-t}}e^{-t}(1-tlnt)}{t}$$

It implies that:

$$\psi(1) = -\gamma + ln\left(\lim_{t\to 0}\frac{e^{-t}t}{t^{e^{-t}}e^{-t}(1-tlnt)}\right) = -\gamma + ln\left(\lim_{t\to 0}\frac{1}{t^{e^{-t}-1}(1-tlnt)}\right)$$
$$= -\gamma + ln\left(\frac{1}{\lim_{t\to 0}\left(t^{e^{-t}-1}\right)\lim_{t\to 0}(1-tlnt)}\right)$$

We also have:

$$\lim_{t\to 0}\left(t^{e^{-t}-1}\right) = \lim_{t\to 0} exp\left(\frac{e^{-t}-1}{1/lnt}\right) = exp\left(\lim_{t\to 0}\frac{e^{-t}-1}{1/lnt}\right)$$

(due to transformation with regard to indeterminate form [7])

$$= exp\left(\lim_{t\to 0}\frac{\mathrm{d}(e^{-t}-1)}{\mathrm{d}(1/lnt)}\right) = exp\left(\lim_{t\to 0}\frac{-e^{-t}}{-\frac{1}{t(lnt)^2}}\right)$$

(using L'Hôpital's rule by taking derivatives of both numerator and denominator [7])

$$= exp\left(\lim_{t\to 0} e^{-t}t(lnt)^2\right) = exp\left(\lim_{t\to 0} t(lnt)^2\right) = exp\left(\lim_{t\to 0}\frac{(lnt)^2}{1/t}\right)$$
$$= exp\left(\lim_{t\to 0}\frac{\mathrm{d}((lnt)^2)}{\mathrm{d}(1/t)}\right) = exp\left(\lim_{t\to 0}\frac{2lnt/t}{-1/t^2}\right)$$

(using L'Hôpital's rule by taking derivatives of both numerator and denominator [7])

$$= exp\left(-\lim_{t\to 0} 2tlnt\right) = exp\left(-\lim_{t\to 0}\frac{2lnt}{1/t}\right)$$
$$= exp\left(-\lim_{t\to 0}\frac{\mathrm{d}(2lnt)}{\mathrm{d}(1/t)}\right) = exp\left(\lim_{t\to 0}\frac{2/t}{1/t^2}\right)$$

(using L'Hôpital's rule by taking derivatives of both numerator and denominator [7])

$$= exp\left(\lim_{t\to 0} 2t\right) = exp(0) = 1$$

We also have:

$$\lim_{t\to 0}(1 - tlnt) = 1 - \lim_{t\to 0}(tlnt) = 1 - \lim_{t\to 0}\frac{lnt}{1/t}$$
$$= 1 - \lim_{t\to 0}\frac{\mathrm{d}(lnt)}{\mathrm{d}(1/t)} = 1 + \lim_{t\to 0}\frac{1/t}{1/t^2}$$

(using L'Hôpital's rule by taking derivatives of both numerator and denominator [7])

$$= 1 + \lim_{t\to 0} t = 1 + 0 = 1$$

Therefore, it implies that:

$$\psi(1) = -\gamma + ln\left(\frac{1}{\lim_{t\to 0}\left(t^{e^{-t}-1}\right)\lim_{t\to 0}(1 - tlnt)}\right) = -\gamma + ln\left(\frac{1}{1 * 1}\right) = -\gamma$$

Briefly, the value $\psi(1)$ is always equal to $-\gamma$ [6]. Given x is positive integer, equation 7 is replaced by equation 9 for calculating digamma function in case of positive integer number.

$$\begin{aligned}&\psi(1) = -\gamma\\ &\psi(x) = -\gamma + \sum_{k=1}^{x-1}\frac{1}{k}\\ &(x \text{ positive integer and } x \geq 2)\end{aligned} \tag{9}$$

Proof,

$$\forall x \geq 2, \psi(x) = \psi(x-1) + \frac{1}{x-1} = \psi(x-2) + \frac{1}{x-2} + \frac{1}{x-1}$$

(by applying equation 8)

$$= \psi(x-3) - \frac{1}{x+3} - \frac{1}{x+2} - \frac{1}{x+1}$$
$$= \cdots = \psi\big(x-(x-1)\big) + \underbrace{\frac{1}{x-(x-1)} + \frac{1}{x-(x-2)} + \cdots + \frac{1}{x-1}}_{x-1}$$
$$= \psi\big(x-(x-1)\big) + \underbrace{\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{x-1}}_{x-1}$$
$$= \psi(1) + \sum_{k=1}^{x-1}\frac{1}{k} = -\gamma + \sum_{k=1}^{x-1}\frac{1}{k}$$

Let $\psi_1(x)$ be the first-order of digamma function, we have:

$$\psi_1(x) = \psi'(x) = \frac{\mathrm{d}\left(\int_0^{+\infty}\left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1-e^{-t}}\right)\mathrm{d}t\right)}{\mathrm{d}x} = \int_0^{+\infty} \frac{\mathrm{d}\left(\frac{e^{-t}}{t} - \frac{e^{-xt}}{1-e^{-t}}\right)}{\mathrm{d}x}\mathrm{d}t$$

(because function $\frac{e^{-t}}{t} - \frac{e^{-xt}}{1-e^{-t}}$ is continuous and differentiable in open interval $(0, +\infty)$ with regard to variable x)

$$= \int_0^{+\infty} \frac{\mathrm{d}\left(-\frac{e^{-xt}}{1-e^{-t}}\right)}{\mathrm{d}x}\mathrm{d}t = \int_0^{+\infty} \frac{te^{-xt}}{1-e^{-t}}\mathrm{d}t$$

We also have:

$$\psi_1(x+y) = \frac{\mathrm{d}\psi(x+y)}{\mathrm{d}x} = \frac{\mathrm{d}\psi(x+y)}{\mathrm{d}y} = \int_0^{+\infty} \frac{\mathrm{d}\left(-\frac{e^{-(x+y)t}}{1-e^{-t}}\right)}{\mathrm{d}x}\mathrm{d}t = \int_0^{+\infty} \frac{\mathrm{d}\left(-\frac{e^{-(x+y)t}}{1-e^{-t}}\right)}{\mathrm{d}y}\mathrm{d}t = \int_0^{+\infty} \frac{te^{-(x+y)t}}{1-e^{-t}}\mathrm{d}t$$

Function $\psi_1(x)$ is also called *trigamma* function; please refer to documents [9], [10] and [11] for more details about trigamma function. Briefly, equation 10 expresses trigamma function [10].

$$\begin{aligned}\psi_1(x) &= \int_0^{+\infty} \frac{te^{-xt}}{1-e^{-t}}\mathrm{d}t \\ \psi_1(x+y) &= \int_0^{+\infty} \frac{te^{-(x+y)t}}{1-e^{-t}}\mathrm{d}t\end{aligned} \tag{10}$$

Suppose variable x is non-zero, we have:

$$\psi_1(x+1) = \int_0^{+\infty} \frac{te^{-(x+1)t}}{1-e^{-t}}\mathrm{d}t = \int_0^{+\infty}\left(\frac{te^{-xt}}{1-e^{-t}} - te^{-xt}\right)\mathrm{d}t$$

$$= \int_0^{+\infty} \frac{te^{-xt}}{1-e^{-t}}\mathrm{d}t - \int_0^{+\infty} te^{-xt}\mathrm{d}t = \psi_1(x) + \frac{1}{x}\int_0^{+\infty} td(e^{-xt})\mathrm{d}t$$

$$= \psi_1(x) + \frac{1}{x}\, te^{-xt}\Big|_0^{+\infty} - \frac{1}{x}\int_0^{+\infty} e^{-xt}\mathrm{d}t$$

$$= \psi_1(x) + \frac{1}{x}\, te^{-xt}\Big|_0^{+\infty} + \frac{1}{x^2}e^{-xt}\Big|_0^{+\infty}$$

$$= \psi_1(x) + \frac{1}{x}\lim_{t\to+\infty}(te^{-xt}) - \frac{1}{x}\lim_{t\to 0}(te^{-xt}) + \frac{1}{x^2}\lim_{t\to+\infty}e^{-xt} - \frac{1}{x^2}\lim_{t\to 0}e^{-xt}$$

$$= \psi_1(x) + \frac{1}{x}\lim_{t\to+\infty}(te^{-xt}) - \frac{1}{x^2} = \psi_1(x) - \frac{1}{x^2} + \frac{1}{x}\lim_{t\to+\infty}\frac{t}{e^{xt}}$$

$$= \psi_1(x) - \frac{1}{x^2} + \frac{1}{x}\lim_{t\to+\infty}\frac{\mathrm{d}(t)}{\mathrm{d}(e^{xt})} = \psi_1(x) - \frac{1}{x^2} + \frac{1}{x}\lim_{t\to+\infty}\frac{1}{xe^{xt}}$$

(using L'Hôpital's rule by taking derivatives of both numerator and denominator [7])

$$= \psi_1(x) - \frac{1}{x^2}$$

Briefly, the recurrence equation of trigamma function is specified by equation 11 [12].

$$\psi_1(x+1) = \psi_1(x) - \frac{1}{x^2} \tag{11}$$

Equation 11 shows recurrence relation [12] of trigamma function, which implicates that it is very easy to compute $\psi_1(x)$ if variable x is positive integer. Thus, it is necessary to calculate $\psi_1(1)$ which is the evaluation of trigamma function at starting positive point 1, we have:

$$\psi_1(x) = \psi'(x) = \frac{d}{dx}\left(\frac{\Gamma'(x)}{\Gamma(x)}\right)$$
$$= \frac{\Gamma''(x)\Gamma(x) - \Gamma'(x)\Gamma'(x)}{\Gamma(x)\Gamma(x)} = \frac{\Gamma''(x)}{\Gamma(x)} - \left(\frac{\Gamma'(x)}{\Gamma(x)}\right)^2$$
$$= \frac{\Gamma''(x)}{\Gamma(x)} - \left(\psi(x)\right)^2$$
$$\Rightarrow \psi_1(x) = \frac{\Gamma''(1)}{\Gamma(1)} - \left(\psi(1)\right)^2 = \frac{\Gamma''(1)}{\Gamma(1)} - \gamma^2$$

We have:

$$\Gamma(1) = \int_0^{+\infty} e^{-t}\mathrm{d}t = -e^{-t}\Big|_0^{+\infty} = \lim_{t\to+\infty}(-e^{-t}) + 1 = 0 + 1 = 1$$

$$\Gamma'(x) = \frac{\mathrm{d}\left(\int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t\right)}{\mathrm{d}x} = \int_0^{+\infty} t^{x-1}e^{-t}lnt\mathrm{d}t$$
$$\Rightarrow \Gamma''(x) = \frac{\mathrm{d}\left(\Gamma'(x)\right)}{\mathrm{d}x} = \frac{\mathrm{d}\left(\int_0^{+\infty} t^{x-1}e^{-t}lnt\mathrm{d}t\right)}{\mathrm{d}x} = \int_0^{+\infty} t^{x-1}e^{-t}(lnt)^2\mathrm{d}t$$
$$\Rightarrow \Gamma''(1) = \int_0^{+\infty} e^{-t}(lnt)^2\mathrm{d}t = \gamma^2 + \frac{\pi^2}{6}$$

Where $\gamma \approx 0.577215$ is Euler-Mascheroni constant [8]. The evaluation $\int_0^{+\infty} e^{-t}(lnt)^2dt = \gamma^2 + \frac{\pi^2}{6}$ is found out in [8].

It implies that

$$\psi_1(1) = \frac{\Gamma''(1)}{\Gamma(1)} - \gamma^2 = \frac{\gamma^2 + \frac{\pi^2}{6}}{1} - \gamma^2 = \frac{\pi^2}{6}$$

Briefly, the value $\psi_1(1)$ is always equal to $\frac{\pi^2}{6}$ [12]. Given x is positive integer, equation 10 is replaced by equation 12 for calculating trigamma function in case of positive integer number, as follows:

$$\psi_1(1) = \frac{\pi^2}{6}$$
$$\psi_1(x) = \frac{\pi^2}{6} - \sum_{k=1}^{x-1}\frac{1}{k^2} \tag{12}$$
$$(x \text{ positive integer and } x \geq 2)$$

Proof,

$$\forall x \geq 2, \psi_1(x) = \psi_1(x-1) - \frac{1}{(x-1)^2} = \psi_1(x-2) - \left(\frac{1}{(x-2)^2} + \frac{1}{(x-1)^2}\right)$$

(by applying equation 11)

$$= \psi_1(x-3) - \left(\frac{1}{(x-3)^2} + \frac{1}{(x-2)^2} + \frac{1}{(x-1)^2}\right)$$

$$= \cdots$$

$$= \psi_1\big(x-(x-1)\big) - \Bigg(\underbrace{\frac{1}{\big(x-(x-1)\big)^2} + \frac{1}{\big(x-(x-2)\big)^2} + \cdots + \frac{1}{(x-1)^2}}_{x-1}\Bigg)$$

$$== \psi_1\big(x-(x-1)\big) - \Bigg(\underbrace{\frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{(x-1)^2}}_{x-1}\Bigg)$$

$$= \psi_1(1) - \sum_{k=1}^{x-1} \frac{1}{k^2} = \frac{\pi^2}{6} - \sum_{k=1}^{x-1} \frac{1}{k^2}$$

In general, I discover two equations 9 and 12 in order to calculate digamma function and trigamma function with regard to positive variable.

The beta function [13] denoted $B(x, y)$ is a special function defined as below:

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{13}$$

Please distinguish beta density function $\beta(X;\ a,\ b)$ specified in equation 5 known as probability density function (PDF) from beta function $B(x, y)$ specified by equation 13.

The first-order partial derivative of $B(x, y)$ is determined as follows:

$$\frac{\partial B(x,y)}{\partial x} = \Gamma(y)\left(\frac{\Gamma'(x)\Gamma(x+y) - \Gamma(x)\Gamma'(x+y)}{\big(\Gamma(x+y)\big)^2}\right) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \frac{\Gamma'(x)\Gamma(x+y) - \Gamma(x)\Gamma'(x+y)}{\Gamma(x)\Gamma(x+y)}$$

$$= B(x,y)\frac{\Gamma'(x)\Gamma(x+y) - \Gamma(x)\Gamma'(x+y)}{\Gamma(x)\Gamma(x+y)} = B(x,y)\left(\frac{\Gamma'(x)}{\Gamma(x)} - \frac{\Gamma'(x+y)}{\Gamma(x+y)}\right)$$

Due to digamma function $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$, we have:

$$\frac{\partial B(x,y)}{\partial x} = B(x,y)\big(\psi(x) - \psi(x+y)\big) \tag{14}$$

The digamma function is always determined by equations 7 and 9. Substituting beta function $B(x,\ y)$ specified equation 13 into equation 5, the beta density function is re-written:

$$f(X;a,b) = \beta(X;a,b) = \frac{1}{B(a,b)} X^{a-1}(1-X)^{b-1} \tag{15}$$

Now we specify the likelihood function of beta distribution by applying equation 15 as below:

$$L(a,b)=\prod_{i=1}^{n}\beta(x_i;a,b)=\prod_{i=1}^{n}\frac{1}{B(a,b)}x_i^{a-1}(1-x_i)^{b-1}=\frac{1}{B^n(a,b)}\left(\prod_{i=1}^{n}x_i^{a-1}\right)\left(\prod_{i=1}^{n}(1-x_i)^{b-1}\right)$$

Taking the logarithm of $L(a, b)$, we have the log-likelihood function for beta distribution as follows:

$$\begin{aligned}LnL(a,b)&=nln\left(\frac{1}{B(a,b)}\right)+\sum_{i=1}^{n}\big((a-1)ln(x_i)\big)+\sum_{i=1}^{n}\big((b-1)ln(1-x_i)\big)\\&=-nln\big(B(a,b)\big)+(a-1)\sum_{i=1}^{n}ln(x_i)+(b-1)\sum_{i=1}^{n}ln(1-x_i)\end{aligned}\tag{16}$$

Fig. 2 [15] shows the graph of log-likelihood function specified by equation 16 with regard to variables a and b given x_1=0.1 and x_2=0.2.

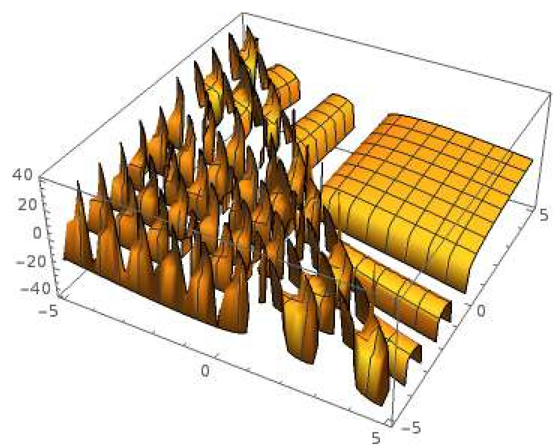

Fig. 2. Log-likelihood function with regard to variables *a* and *b*

Fig. 3 [15] shows the contour of log-likelihood function specified by equation 16 with regard to variables a and b given x_1=0.1 and x_2=0.2.

Please pay attention to equation 16 because equation 16 is specific case of equation 3 mentioned in previous section 2; thus, MLE is applied into beta distribution.

Note that $LnL(a, b) = -\infty$ if any instance x_i is equal to 1 or 0. In practice, we should assign a very large number to $LnL(a, b)$ in this case, instead of keeping the infinity.

Two parameters a and b must be determined so that they maximize the log-likelihood function. Thus, by taking two first-order partial derivatives of log-likelihood function specified in equation 16 corresponding to two parameters and by applying equation 14, we have:

$$\frac{\partial LnL(a,b)}{\partial a}=-n\frac{1}{B(a,b)}\frac{\partial B(a,b)}{\partial a}+\sum_{i=1}^{n}ln(x_i)=-n\big(\psi(a)-\psi(a+b)\big)+\sum_{i=1}^{n}ln(x_i)\tag{17}$$

$$\frac{\partial LnL(a,b)}{\partial b} = -n\frac{1}{B(a,b)}\frac{\partial B(a,b)}{\partial b} + \sum_{i=1}^{n} ln(1-x_i)$$
$$= -n\big(\psi(b) - \psi(a+b)\big) + \sum_{i=1}^{n} ln(1-x_i) \tag{18}$$

Where $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is digamma function specified by equations 7 and 9. Note that notation $\frac{\partial f}{\partial x}$ denotes first-order partial derivative of multi-variable function f with regard to variable x.

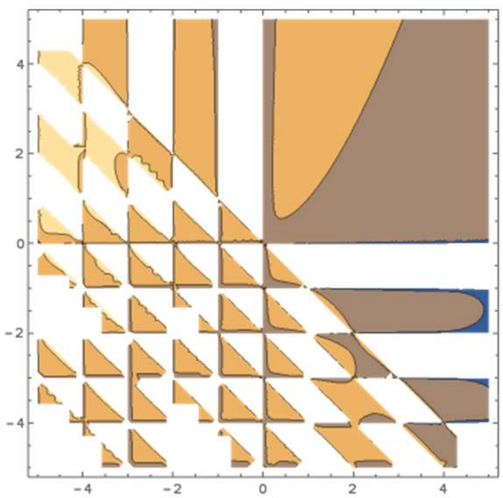

Fig. 3. Contour of log-likelihood function with regard to variables *a* and *b*

Please pay attention to equations 17 and 18 for determining two first-order partial derivatives of log-likelihood function of beta distribution. Setting such two partial derivatives equal 0 so as to find out two parameters *a* and *b*, we have a set of equations whose two solutions are the values of *a* and *b* [14, p. 288]:

$$\begin{cases} \frac{\partial LnL(a,b)}{\partial a} = 0 \\ \frac{\partial LnL(a,b)}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} \psi(a) - \psi(a+b) = \frac{1}{n}\sum_{i=1}^{n} ln(x_i) \\ \psi(b) - \psi(a+b) = \frac{1}{n}\sum_{i=1}^{n} ln(1-x_i) \end{cases} \tag{19}$$

Equation 19 shows the set of differential equations for estimating parameters *a* and *b*. Author [14] proposes an algorithm to find out the approximate solutions. Such algorithm will be mentioned in next section.

According to equation 9, given *a* and *b* are positive integers, the digamma function $\psi(x)$ is:

$$\psi(x) = -\gamma + \sum_{k=1}^{x-1} \frac{1}{k}$$

We have:

$$\begin{cases} \psi(a) - \psi(a+b) = \dfrac{1}{n}\sum_{i=1}^{n} ln(x_i) \\ \psi(b) - \psi(a+b) = \dfrac{1}{n}\sum_{i=1}^{n} ln(1-x_i) \end{cases} \Leftrightarrow \begin{cases} n\big(\psi(a) - \psi(a+b)\big) = \sum_{i=1}^{n} ln(x_i) \\ n\big(\psi(b) - \psi(a+b)\big) = \sum_{i=1}^{n} ln(1-x_i) \end{cases}$$

$$\Leftrightarrow \begin{cases} n\left(-\gamma + \sum_{k=1}^{a-1}\dfrac{1}{k} + \gamma - \sum_{k=1}^{a+b-1}\dfrac{1}{k}\right) = \sum_{i=1}^{n} ln(x_i) \\ n\left(-\gamma + \sum_{k=1}^{b-1}\dfrac{1}{k} + \gamma - \sum_{k=1}^{a+b-1}\dfrac{1}{k}\right) = \sum_{i=1}^{n} ln(1-x_i) \end{cases}$$

$$\Leftrightarrow \begin{cases} -n\sum_{k=a}^{a+b-1}\dfrac{1}{k} = \sum_{i=1}^{n} ln(x_i) \\ -n\sum_{k=b}^{a+b-1}\dfrac{1}{k} = \sum_{i=1}^{n} ln(1-x_i) \end{cases} \Rightarrow \begin{cases} exp\left(-n\sum_{k=a}^{a+b-1}\dfrac{1}{k}\right) = \prod_{i=1}^{n} x_i \\ exp\left(-n\sum_{k=b}^{a+b-1}\dfrac{1}{k}\right) = \prod_{i=1}^{n}(1-x_i) \end{cases}$$

(By taking exponent function of both sides of these equations)

Briefly, the parameter estimators $\hat{a}$ and $\hat{b}$ are solutions of two following equations specified by equation 20 in case that *a* and *b* are positive integer numbers.

$$\begin{cases} exp\left(-n\sum_{k=a}^{a+b-1}\dfrac{1}{k}\right) = \prod_{i=1}^{n} x_i \\ exp\left(-n\sum_{k=b}^{a+b-1}\dfrac{1}{k}\right) = \prod_{i=1}^{n}(1-x_i) \end{cases} \Leftrightarrow \begin{cases} G_1(a,b) = L_1 \\ G_2(a,b) = L_2 \end{cases} \tag{20}$$

Where,

$$0 \le x \le 1$$

$$G_1(a,b) = exp\left(-n\sum_{k=a}^{a+b-1}\frac{1}{k}\right) \text{ and } G_2(a,b) = exp\left(-n\sum_{k=b}^{a+b-1}\frac{1}{k}\right)$$

$$L_1 = \prod_{i=1}^{n} x_i \text{ and } L_2 = \prod_{i=1}^{n}(1-x_i)$$

Next section will illustrates how to solve equation 20. Now it is necessary to research the co-variance matrix *Var*(*a*, *b*) of parameters of beta density function mentioned in previous section. Let *H*(*a*, *b*) be the second-order partial derivative matrix called Hessian matrix, we have:

$$H(a,b) = \begin{pmatrix} \dfrac{\partial^2 LnL}{\partial a^2} & \dfrac{\partial^2 LnL}{\partial a \partial b} \\ \dfrac{\partial^2 LnL}{\partial b \partial a} & \dfrac{\partial^2 LnL}{\partial b^2} \end{pmatrix}$$

Note, the bracket (.) denotes matrix.

Basing on equations 17 and 18, we can determine four second-order partial derivatives of log-likelihood function as follows:

$$\frac{\partial^2 LnL}{\partial a^2} = \frac{\partial \psi(a+b)}{\partial a} - n\frac{\partial \psi(a)}{\partial a} = \psi_1(a+b) - n\psi_1(a)$$

$$\frac{\partial^2 LnL}{\partial a \partial b} = \frac{\partial \psi(a+b)}{\partial b} = \psi_1(a+b)$$

$$\frac{\partial^2 LnL}{\partial b \partial a} = \frac{\partial \psi(a+b)}{\partial a} = \psi_1(a+b)$$

$$\frac{\partial^2 LnL}{\partial b^2} = \frac{\partial \psi(a+b)}{\partial b} - n\frac{\partial \psi(b)}{\partial b} = \psi_1(a+b) - n\psi_1(b)$$

Where $\psi_1(.)$ denotes trigamma function specified by equations 10 and 12. According to equation 4, the co-variance matrix $Var(a, b)$ is the inversion of negative expectation of Hessian matrix. Please read the book "Linear Algebra" by author [16, p. 134] and the book "Linear Algebra and Its Applications" by author [17, pp. 102-109] for more details of how to take inversion of a given matrix. We have:

$$\begin{aligned} Var(a,b) &= \left(-E\big(H(a,b)\big)\right)^{-1} \\ &= \left(-E\begin{pmatrix}\psi_1(a+b) - n\psi_1(a) & \psi_1(a+b) \\ \psi_1(a+b) & \psi_1(a+b) - n\psi_1(b)\end{pmatrix}\right)^{-1} \\ &= \left(-\begin{pmatrix}\psi_1(a+b) - n\psi_1(a) & \psi_1(a+b) \\ \psi_1(a+b) & \psi_1(a+b) - n\psi_1(b)\end{pmatrix}\right)^{-1} \end{aligned}$$

(Because trigamma functions $\psi_1(a)$, $\psi_1(b)$, and $\psi_1(a+b)$ are only dependent on parameters a and b, the expectation of $H(a, b)$ is merely $H(a, b)$)

$$\begin{aligned} &= \begin{pmatrix}n\psi_1(a) - \psi_1(a+b) & -\psi_1(a+b) \\ -\psi_1(a+b) & n\psi_1(b) - \psi_1(a+b)\end{pmatrix}^{-1} \\ &= \frac{1}{n^2\psi_1(a)\psi_1(b) - n\psi_1(a+b)\big(\psi_1(a)+\psi_1(b)\big)} * \begin{pmatrix}n\psi_1(b) - \psi_1(a+b) & \psi_1(a+b) \\ \psi_1(a+b) & n\psi_1(a) - \psi_1(a+b)\end{pmatrix} \end{aligned}$$

Briefly, equation 21 specifies the co-variance matrix of parameters of beta density function as follows:

$$Var(a,b) = A\begin{pmatrix}n\psi_1(b) - \psi_1(a+b) & \psi_1(a+b) \\ \psi_1(a+b) & n\psi_1(a) - \psi_1(a+b)\end{pmatrix} \quad (21)$$

Where $\psi_1(.)$ denotes trigamma function and,

$$A = \frac{1}{n^2\psi_1(a)\psi_1(b) - n\psi_1(a+b)\big(\psi_1(a)+\psi_1(b)\big)}$$

Equation 21 is concrete case of equation 4 when probability distribution is beta distribution. If parameters a and b are positive integers, the trigamma function $\psi_1(.)$ is calculated simply according to equation 12; this is the ultimate purpose of this section.

The roots of diagonal elements are the standard deviations (standard errors) of parameter estimates. Let $\sigma(\hat{a})$ and $\sigma(\hat{b})$ be the standard errors of optimal parameters $\hat{a}$ and $\hat{b}$ where $\hat{a}$ and $\hat{b}$ are solutions of equations specified by equation 20, we have:

$$\begin{aligned}\sigma(\hat{a}) &= \sqrt{A\left(n\psi_1(\hat{b}) - \psi_1(\hat{a}+\hat{b})\right)}\\ \sigma(\hat{b}) &= \sqrt{A\left(n\psi_1(\hat{a}) - \psi_1(\hat{a}+\hat{b})\right)}\end{aligned} \tag{22}$$

Where $\psi_1(.)$ denotes the trigamma function and,

$$A = \frac{1}{n^2\psi_1(a)\psi_1(b) - n\psi_1(a+b)\big(\psi_1(a)+\psi_1(b)\big)}$$

Equation 22 specifying standard errors of parameter estimates ends up this section mentioning applying MLE technique into beta distribution. Now the next section is an example of beta likelihood estimation.

3 An Application of Beta Likelihood Estimation to Specify Prior Probabilities in Bayesian Network

Recall that the parameter estimators $\hat{a}$ and $\hat{b}$ are solutions of two equations, according to equation 20 as follows:

$$\begin{cases} exp\left(-n\displaystyle\sum_{k=a}^{a+b-1}\frac{1}{k}\right) = \displaystyle\prod_{i=1}^{n} x_i \\ exp\left(-n\displaystyle\sum_{k=b}^{a+b-1}\frac{1}{k}\right) = \displaystyle\prod_{i=1}^{n}(1-x_i)\end{cases} \Leftrightarrow \begin{cases} G_1(a,b) = L_1 \\ G_2(a,b) = L_2\end{cases}$$

Where,

$$G_1(a,b) = exp\left(-n\sum_{k=a}^{a+b-1}\frac{1}{k}\right) \text{ and } G_2(a,b) = exp\left(-n\sum_{k=b}^{a+b-1}\frac{1}{k}\right)$$

$$L_1 = \prod_{i=1}^{n} x_i \text{ and } L_2 = \prod_{i=1}^{n}(1-x_i)$$

Author [14] proposes the iterative algorithm that each pair values (a_i, b_i) which are values of variables a and b are fed to G_1, G_2 at each iteration. Two biases $\Delta_1=G_1(a_i, b_i)-L_1$ and $\Delta_2=G_2(a_i, b_i)-L_2$ are computed. The normal bias is the root of sum of the second power Δ_1 and the second power of Δ_2 and so we have $\Delta=\sqrt{\Delta_1^2+\Delta_2^2}$. The pair $(\hat{a}, \hat{b})$ whose normal bias Δ is minimum are chosen as the parameter estimators. The algorithm is described in Table 1 as below [14, p. 291]:

Table 1. Iterative algorithm to estimate parameters *a* and *b*

```
minΔ = +∞
â = b̂ = 1  (uniform distribution )
For a=1 to n do
    For b=1 to n do
        Δ1=G1(a, b)–L1
        Δ2=G2(a, b)–L2
        Δ=sqrt(Δ1^2 + Δ2^2)
        If Δ < minΔ then
            minΔ=Δ
            â=a
            b̂=b
        End If
    End For a
End For b
(â and b̂ are optimal parameters)
```

Where *min* Δ denotes minimum bias.

The main application of beta likelihood estimation is to specify prior probabilities of Bayesian network. Bayesian network (BN) is a directed acyclic graph constituted of a set of nodes representing random variables and a set of arcs representing relationships among nodes. In general, BN consists of qualitative model quantitative model. The qualitative model is its structure and the quantitative model is its parameters, namely conditional probability tables (CPT) whose entries are probabilities quantifying relationships among variables. For example, there is a BN having two binary variables X_1, X_2 and one arc which links them together in which X_2 is conditional dependent on X_1. Each variable X_i owns a CPT. Fig. 4 is an example of BN with two nodes X_1 and X_2 whose CPT (s) are not specified yet.

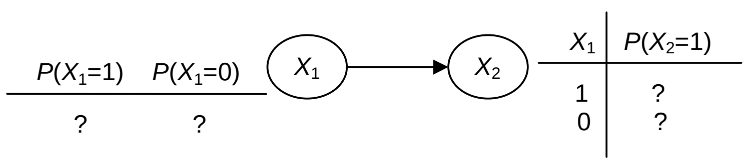

Fig. 4. Bayesian network in which CPT (s) are not specified yet

CPT (s) are parameters of BN. The quality of CPT depends on the initialized values of its entries. Such initial values are prior probabilities. If prior probabilities are already specified, the expectation maximization (EM) algorithm can be used to improve them even in case of missing data [18, pp. 359-363]. However, this research focuses on applying beta likelihood estimation aforementioned in previous section into specifying the prior probabilities. Your attention please, both EM algorithm and beta likelihood estimation are parameter learning methods. Beta distribution is often used to represent CPT. Let $\beta_1(a_1,b_1)$, $\beta_2(a_2,b_2)$, $\beta_3(a_3,b_3)$ be beta distributions of conditional probabilities $P(X_1=1)$, $P(X_2=1|X_1=1)$ and $P(X_2=1|X_1=0)$. These probabilities are expectations of beta distribution [18, p. 302].

$$P(X_1 = 1) = E\big(\beta_1(a_1, b_1)\big) = \frac{a_1}{a_1 + b_1}$$
$$P(X_2 = 1 \mid X_1 = 1) = E\big(\beta_2(a_2, b_2)\big) = \frac{a_2}{a_2 + b_2}$$
$$P(X_2 = 1 \mid X_1 = 0) = E\big(\beta_2(a_3, b_3)\big) = \frac{a_3}{a_3 + b_3}$$

It is necessary to determine three parameter pairs (a_1, b_1), (a_2, b_2) and (a_3, b_3) of three beta distributions β_1, β_2 and β_3, respectively in order to specify prior probabilities $P(X_1=1)$, $P(X_2=1|X_1=1)$ and $P(X_2=1|X_1=0)$. Suppose we perform 5 trials of a random process, the outcome of i^{th} trial denoted $D^{(i)}$ is considered as an evidence in which X_1 and X_2 obtain value 0 or 1. So we have the vector of 5 evidences $\mathcal{D} = (D^{(1)}, D^{(2)}, D^{(2)}, D^{(3)}, D^{(4)}, D^{(5)})$. Table 2 shows these evidences.

Table 2. The evidences corresponding to 5 trials

	X_1	X_2
$D^{(1)}$	$X_1 = 1$	$X_2 = 1$
$D^{(2)}$	$X_1 = 1$	$X_2 = 1$
$D^{(3)}$	$X_1 = 1$	$X_2 = 1$
$D^{(4)}$	$X_1 = 1$	$X_2 = 0$
$D^{(5)}$	$X_1 = 0$	$X_2 = 0$

According to the algorithm described in Table 1, let L_{ij}, G_{ij}, Δ_{ij}, Δ_i be the values of L_j, G_j, Δ_j, Δ with respect to β_i where $i = \overline{1,3}$ and $j = \overline{1,2}$. We have:

- $L_{11}=\prod_{i=1}^{n} x_i$ and $L_{12}=\prod_{i=1}^{n}(1 - x_i)$ where x_i is the instance of X_1.
- $L_{21}=\prod_{i=1}^{n} x_i$ and $L_{22}=\prod_{i=1}^{n}(1 - x_i)$ where x_i is the instance of X_2 given $X_1=1$.
- $L_{31}=\prod_{i=1}^{n} x_i$ and $L_{32}=\prod_{i=1}^{n}(1 - x_i)$ where x_i is the instance of X_2 given $X_1=0$.
- $G_{11}(a_1, b_1)=exp\left(-n\sum_{k=a_1}^{a_1+b_1-1}\frac{1}{k}\right)$ and $G_{12}(a_1, b_1)=exp\left(-n\sum_{k=b_1}^{a_1+b_1-1}\frac{1}{k}\right)$
- $G_{21}(a_2, b_2)=exp\left(-n\sum_{k=a_2}^{a_2+b_2-1}\frac{1}{k}\right)$ and $G_{22}(a_2, b_2)=exp\left(-n\sum_{k=b_2}^{a_2+b_2-1}\frac{1}{k}\right)$
- $G_{31}(a_3, b_3)=exp\left(-n\sum_{k=a_3}^{a_3+b_3-1}\frac{1}{k}\right)$ and $G_{32}(a_3, b_3)=exp\left(-n\sum_{k=b_3}^{a_3+b_3-1}\frac{1}{k}\right)$
- $\Delta_{11}=G_{11} - L_{11}$, $\Delta_{12}=G_{12} - L_{12}$ and $\Delta_1=\sqrt{\Delta_{11}^2 + \Delta_{12}^2}$
- $\Delta_{21}=G_{21} - L_{21}$, $\Delta_{22}=G_{22} - L_{22}$ and $\Delta_2=\sqrt{\Delta_{21}^2 + \Delta_{22}^2}$
- $\Delta_{31}=G_{31} - L_{31}$, $\Delta_{32}=G_{32} - L_{32}$ and $\Delta_3=\sqrt{\Delta_{31}^2 + \Delta_{32}^2}$

Let D_1, D_2, and D_3 be the set of x_i (s) that are instances of X_1, X_2 given $X_1=1$, and X_2 given $X_1=0$, respectively. From evidences expressed in Table 2, we have:

$$D_1 = \{x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 1, x_5 = 0\}$$
$$D_2 = \{x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 0\}$$
$$D_3 = \{x_1 = 0\}$$

Each instance x_i will be modified so that products $\prod_{i=1}^{n} x_i$ and $\prod_{i=1}^{n}(1 - x_i)$ avoid getting zero frequently.

- If x_i equals 1, it is subtracted by a very small number ε, for example, given $\varepsilon=0.1$, $x_i = x_i-0.1 = 1-0.1 = 0.9$.
- If x_i equals 0, it is added by a very small number ε, for example, $\varepsilon=0.1$, $x_i = x_i+0.1 = 0+0.1 = 0.1$.

Thus, we have:

$$D_1 = \{x_1 = 0.9, x_2 = 0.9, x_3 = 0.9, x_4 = 0.9, x_5 = 0.1\}$$
$$D_2 = \{x_1 = 0.9, x_2 = 0.9, x_3 = 0.9, x_4 = 0.1\}$$
$$D_3 = \{x_1 = 0.1\}$$

Suppose the range of all parameters is from 1 to 4. By applying the algorithm described in Table 1, it is easy to compute the normal biases. For example, given $a_1 = 1$ and $b_1 = 1$, we have:

$$L_{11}(a_1 = 1, b_1 = 1) = \prod_{x_i \in D_1} x_i = 0.9 * 0.9 * 0.9 * 0.9 * 0.1 \approx 0.0656$$

$$L_{12}(a_1 = 1, b_1 = 1) = \prod_{x_i \in D_1} (1 - x_i) = (1 - 0.9) * (1 - 0.9) * (1 - 0.9) * (1 - 0.9) * (1 - 0.1) \approx 0.0001$$

$$G_{11}(a_1 = 1, b_1 = 1) = exp\left(-5 \sum_{k=1}^{1+1-1} \frac{1}{k}\right) = exp\left(-5 * \frac{1}{1}\right) \approx 0.0067$$

$$G_{12}(a_1 = 1, b_1 = 1) = exp\left(-5 \sum_{k=1}^{1+1-1} \frac{1}{k}\right) = exp\left(-5 * \frac{1}{1}\right) \approx 0.0067$$

$$\Delta_{11} = G_{11} - L_{11} = 0.0067 - 0.0656 \approx -0.0589$$

$$\Delta_{12} = G_{12} - L_{12} = 0.0067 - 0.0001 \approx 0.0066$$

$$\Delta_1 = \sqrt{\Delta_{11}^2 + \Delta_{12}^2} = \sqrt{(-0.0589)^2 + (0.0066)^2} \approx 0.0592$$

Table 3 shows normal biases of all possible values of (a_1, b_1).

The normal biases of all possible values of (a_2, b_2) with respect to β_2 are shown in Table 4.

The normal biases of all possible values of (a_3, b_3) with respect to β_3 are shown in Table 5.

Table 3. The normal biases of (a_1, b_1) with respect to β_1

a_1	b_1	L_{11}	L_{12}	G_{11}	G_{12}	Δ_{11}	Δ_{12}	Δ_1
1	1	0.0656	0.0001	0.0067	0.0067	–0.0589	0.0066	0.0592
1	2	0.0656	0.0001	0.0006	0.0821	–0.0651	0.0820	0.1047
1	3	0.0656	0.0001	0.0001	0.1889	–0.0655	0.1888	0.1998
1	4	0.0656	0.0001	0.0000	0.2865	–0.0656	0.2864	0.2938
2	1	0.0656	0.0001	0.0821	0.0006	0.0165	0.0005	0.0165
2	2	0.0656	0.0001	0.0155	0.0155	–0.0501	0.0154	0.0524
2	3	0.0656	0.0001	0.0044	0.0541	–0.0612	0.0540	0.0816
2	4	0.0656	0.0001	0.0016	0.1054	–0.0640	0.1053	0.1232
3	1	0.0656	0.0001	0.1889	0.0001	0.1233	0.0000	0.1233
3	**2**	**0.0656**	**0.0001**	**0.0541**	**0.0044**	**–0.0115**	**0.0044**	**0.0123**
3	3	0.0656	0.0001	0.0199	0.0199	–0.0457	0.0198	0.0498
3	4	0.0656	0.0001	0.0087	0.0458	–0.0570	0.0457	0.0730
4	1	0.0656	0.0001	0.2865	0.0000	0.2209	–0.0001	0.2209
4	2	0.0656	0.0001	0.1054	0.0016	0.0398	0.0015	0.0398
4	3	0.0656	0.0001	0.0458	0.0087	–0.0198	0.0086	0.0216
4	4	0.0656	0.0001	0.0224	0.0224	–0.0432	0.0223	0.0486

From Tables 3, 4, and 5, we recognize that when **(a_1,b_1)=(3,2)**, **(a_2,b_2)=(4,3)**, and **(a_3,b_3)=(1,4)**, the normal biases of distributions β_1, β_2, and β_3, respectively become minimum. So the parameter estimators $(\hat{a}_1, \hat{b}_1)$, $(\hat{a}_2, \hat{b}_2)$, and $(\hat{a}_3, \hat{b}_3)$ corresponding to distributions β_1, β_2, and β_3 are (3,2), (4,3), and (1,4), respectively. So the prior conditional probabilities $P(X_1=1)$, $P(X_2=1|X_1=1)$ and $P(X_2=1|X_1=0)$ are determined:

$$P(X_1 = 1) = \frac{\hat{a}_1}{\hat{a}_1 + \hat{b}_1} = \frac{3}{3 + 2} = 0.6$$

$$P(X_2 = 1 \mid X_1 = 1) = \frac{\hat{a}_2}{\hat{a}_2 + \hat{b}_2} = \frac{4}{4 + 3} \approx 0.57$$

$$P(X_2 = 1 \mid X_1 = 0) = \frac{\hat{a}_3}{\hat{a}_3 + \hat{b}_3} = \frac{1}{1 + 4} = 0.2$$

When these prior probabilities were calculated, the BN is totally determined with full of prior CPT (s) as in Fig. 5.

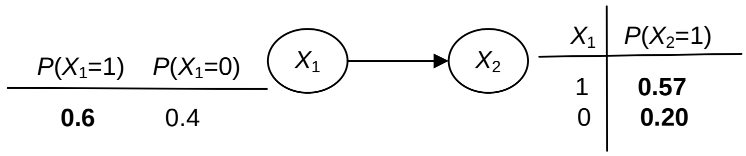

Fig. 5. Bayesian network with full of prior CPT (s)

Table 4. The normal biases of (a_2, b_2) with respect to β_2

a_2	b_2	L_{21}	L_{22}	G_{21}	G_{22}	Δ_{21}	Δ_{22}	Δ_2
1	1	0.0729	0.0009	0.0183	0.0183	–0.0546	0.0174	0.0573
1	2	0.0729	0.0009	0.0025	0.1353	–0.0704	0.1344	0.1518
1	3	0.0729	0.0009	0.0007	0.2636	–0.0722	0.2627	0.2725
1	4	0.0729	0.0009	0.0002	0.3679	–0.0727	0.3670	0.3741
2	1	0.0729	0.0009	0.1353	0.0025	0.0624	0.0016	0.0625
2	2	0.0729	0.0009	0.0357	0.0357	–0.0372	0.0348	0.0509
2	3	0.0729	0.0009	0.0131	0.0970	–0.0598	0.0961	0.1132
2	4	0.0729	0.0009	0.0059	0.1653	–0.0670	0.1644	0.1775
3	1	0.0729	0.0009	0.2636	0.0007	0.1907	–0.0002	0.1907
3	2	0.0729	0.0009	0.0970	0.0131	0.0241	0.0122	0.0270
3	3	0.0729	0.0009	0.0436	0.0436	–0.0293	0.0427	0.0518
3	4	0.0729	0.0009	0.0224	0.0849	–0.0505	0.0840	0.0980
4	1	0.0729	0.0009	0.3679	0.0002	0.2950	–0.0007	0.2950
4	2	0.0729	0.0009	0.1653	0.0059	0.0924	0.0050	0.0925
4	**3**	**0.0729**	**0.0009**	**0.0849**	**0.0224**	**0.0120**	**0.0215**	**0.0246**
4	4	0.0729	0.0009	0.0479	0.0479	–0.0250	0.0470	0.0532

Table 5. The normal biases of (a_3, b_3) with respect to β_3

a_3	b_3	L_{31}	L_{32}	G_{31}	G_{32}	Δ_{31}	Δ_{32}	Δ_3
1	1	0.1	0.9	0.3679	0.3679	0.2679	–0.5321	0.5957
1	2	0.1	0.9	0.2231	0.6065	0.1231	–0.2935	0.3183
1	3	0.1	0.9	0.1599	0.7165	0.0599	–0.1835	0.1930
1	**4**	**0.1**	**0.9**	**0.1245**	**0.7788**	**0.0245**	**–0.1212**	**0.1237**
2	1	0.1	0.9	0.6065	0.2231	0.5065	–0.6769	0.8454
2	2	0.1	0.9	0.4346	0.4346	0.3346	–0.4654	0.5732
2	3	0.1	0.9	0.3385	0.5580	0.2385	–0.3420	0.4169
2	4	0.1	0.9	0.2771	0.6376	0.1771	–0.2624	0.3166
3	1	0.1	0.9	0.7165	0.1599	0.6165	–0.7401	0.9633
3	2	0.1	0.9	0.5580	0.3385	0.4580	–0.5615	0.7246
3	3	0.1	0.9	0.4569	0.4569	0.3569	–0.4431	0.5690
3	4	0.1	0.9	0.3867	0.5397	0.2867	–0.3603	0.4604
4	1	0.1	0.9	0.7788	0.1245	0.6788	–0.7755	1.0306
4	2	0.1	0.9	0.6376	0.2771	0.5376	–0.6229	0.8228
4	3	0.1	0.9	0.5397	0.3867	0.4397	–0.5133	0.6759
4	4	0.1	0.9	0.4679	0.4679	0.3679	–0.4321	0.5675

Let $\sigma(\hat{a}_1)$, $\sigma(\hat{b}_1)$, $\sigma(\hat{a}_2)$, $\sigma(\hat{b}_2)$, $\sigma(\hat{a}_3)$, and $\sigma(\hat{b}_3)$ be standard errors of $\hat{a}_1$, $\hat{b}_1$, $\hat{a}_2$, $\hat{b}_2$, $\hat{a}_3$, and $\hat{b}_3$. By applying equation 22, it is easy to determine these standard errors as follows:

$$A_1 = \frac{1}{5^2\psi_1(\hat{a}_1)\psi_1(\hat{b}_1) - 5\psi_1(\hat{a}_1 + \hat{b}_1)\left(\psi_1(\hat{a}_1) + \psi_1(\hat{b}_1)\right)} = 0.191684$$

$$\sigma(\hat{a}_1) = \sqrt{A_1\left(5\psi_1(\hat{b}_1) - \psi_1(\hat{a}_1 + \hat{b}_1)\right)} = 0.758744$$

$$\sigma(\hat{b}_1) = \sqrt{A_1\left(5\psi_1(\hat{a}_1) - \psi_1(\hat{a}_1 + \hat{b}_1)\right)} = 0.579731$$

$$A_2 = \frac{1}{4^2\psi_1(\hat{a}_2)\psi_1(\hat{b}_2) - 4\psi_1(\hat{a}_2 + \hat{b}_2)\left(\psi_1(\hat{a}_2) + \psi_1(\hat{b}_2)\right)} = 0.726437$$

$$\sigma(\hat{a}_2) = \sqrt{A_2\left(4\psi_1(\hat{b}_2) - \psi_1(\hat{a}_2 + \hat{b}_2)\right)} = 1.01786$$

$$\sigma(\hat{b}_2) = \sqrt{A_2\left(4\psi_1(\hat{a}_2) - \psi_1(\hat{a}_2 + \hat{b}_2)\right)} = 0.844498$$

$$A_3 = \frac{1}{1^2\psi_1(\hat{a}_3)\psi_1(\hat{b}_3) - 1\psi_1(\hat{a}_3 + \hat{b}_3)\left(\psi_1(\hat{a}_3) + \psi_1(\hat{b}_3)\right)} = 25.0051$$

$$\sigma(\hat{a}_3) = \sqrt{A_3\left(1\psi_1(\hat{b}_3) - \psi_1(\hat{a}_3 + \hat{b}_3)\right)} = 1.25013$$

$$\sigma(\hat{b}_3) = \sqrt{A_3\left(1\psi_1(\hat{a}_3) - \psi_1(\hat{a}_3 + \hat{b}_3)\right)} = 5.96637$$

The errors $\sigma(\hat{a}_1)$ and $\sigma(\hat{b}_1)$ are minimum because the number of instances of X_i is 5 – the largest, which implies that $\hat{a}_1$ and $\hat{b}_1$ are best estimates.

In general, the iterative algorithm for solving simple equations specified by equation 20 is the result of applying MLE method into beta density function.

4 Conclusion

This research shares the same methodology with the previous research [14] where positive integer parameters of beta distribution are estimated based on interesting features of gamma function. The ultimate purpose is to simplify solving differential equations in order to estimate such integer parameters by easiest way. The resulted equations are not absolutely simpler than ones from [14] but this research digs deeply into mathematical functions relevant to gamma function such as digamma and trigamma. Consequently, this research is more general and all equations are proved in detail.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Lynch SM. Introduction to applied Bayesian statistics and estimation for social scientists. Springer Berlin Heidelberg NewYork; 2007.

[2] Montgomery DC, Runger GC. Applied statistics and probability for engineers (3rd Edition ed.). New York, NY, USA: John Wiley & Sons, Inc; 2003.

[3] Zivot E. Maximum likelihood estimation. Lecture notes on course "Econometric theory I: Estimation and Inference (first quarter, second year PhD)". University of Washington, Seattle, Washington, USA; 2009.

[4] Johansen I. Graph software. Graph version 4.4.2 build 543(4.4.2 build 543). GNU General Public License; 2012.

[5] Medina LA, Moll VH. The integrals in Gradshteyn and Ryzhik. Part 10: The digamma function. Series A: Mathematical Sciences. 2009;17(2009):45-66.

[6] Wikipedia. Digamma function. (Wikimedia Foundation); 2014.
Available: http://en.wikipedia.org/wiki/Digamma_function
(Retrieved November 9, 2014)

[7] Wikipedia. Indeterminate form. (Wikimedia Foundation); 2014.
Available:http://en.wikipedia.org/wiki/Indeterminate_form
(Retrieved November 11, 2014)

[8] Weisstein EW. (n.d.). Euler-Mascheroni constant. (Wolfram Research); 2014.
Available:http://mathworld.wolfram.com/Euler-MascheroniConstant.html
(Retrieved November 13, 2014)

[9] Weisstein EW. (n.d.). Polygamma function. (Wolfram research); 2014.
Available: http://mathworld.wolfram.com/PolygammaFunction.html
(Retrieved November 13, 2014)

[10] Wikipedia. Polygamma function. (Wikimedia foundation); 2014.
Available:http://en.wikipedia.org/wiki/Digamma_function
(Retrieved November 10, 2014)

[11] Weisstein EW. (n.d.). Trigamma function. (Wolfram research); 2014.
Available: http://mathworld.wolfram.com/TrigammaFunction.html
(Retrieved November 13, 2014)

[12] Wikipedia. Trigamma function. (Wikimedia foundation); 2015.
Available:https://en.wikipedia.org/wiki/Trigamma_function
(Retrieved April 9, 2016)

[13] Wikipedia. Beta function. (Wikimedia foundation); 2014.
Available:Wikipedia website: http://en.wikipedia.org/wiki/Beta_function
(Retrieved November 9, 2014)

[14] Nguyen L. Specifying prior probabilities in Bayesian network by maximum likelihood estimation method. (B. N. Buszewski, Ed.) Sylwan Journal. 2016;(2):281-298.
Available: http://sylwan.ibles.org/syl/index.php/jTracker/index/mWEME

[15] Wolfram. (n.d.). Mathematica. Wolfram mathematica. 2016;10(4). Wolfram research. Available:http://www.wolfram.com/mathematica
(Retrieved March 2016)

[16] Nguyen VH. Linear algebra. Hanoi, Vietnam: Hanoi National University Publishing House; 1999.

[17] Lay DC. Linear algebra and its applications (4th Edition ed.). (D. Lynch, Ed.) Boston, Maryland, USA: Pearson Education; 2012.

[18] Neapolitan RE. Learning Bayesian networks. Upper Saddle River, New Jersey, USA: Prentice Hall; 2003.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/14364