HW6 Biostat 250B

Elvis Cui
Han Cui
Dept. of Biostat
UCLA

1 Suppose X has a Binomial distribution with parameter n and p. Find the posterior density of p if we have a Beta(a,b) prior distribution for p. Calculate the mean and variance of the posterior density and describe their relationships with those of the Beta(a,b). Does it make sense to use method of moment estimators to estimate the mean and variance of this posterior density? Explain.

2 Suppose X has an inverted gamma with density given by $f(x,a,b) = b^a x^{-a-1} \exp(-b/x)/\Gamma(a)$, where a>0 and b>0 and Γ is the Gamma function.

(i) Describe how values of a and b affect the density of X in terms of shape, skewness and symmetry.
(ii) Find its mean and describe how this density is typically used in estimating parameters in a linear model using a Bayesian paradigm.

3 Suppose Z has a density proportional to $z^{n/2-1} \exp(-nz/2)$ and conditional on Z, X|Z=z is multivariate normal with mean 0 and covariance (1/z)I, where I is the k x k identity matrix. What is the density of X? Determine its mean and covariance matrix of X.

4 Ex. 3m, #2 from Seber and Lee's book.

5 Ex. 3m, #3 from Seber and Lee's book.

6 Last week, your supervisor came to you with the paper on Bayes Estimation of Two-Phase Linear Regression Model posted on Week 8 on the class website. She was very interested in the paper and want to apply the methodology in the paper to solve a public health problem. But she cannot understand much of the paper and have no idea how to apply the methodology to a public health problem. Your tasks are (i) explain to her in simple English in not more than ¾ page what the contents and methodology of the paper are about, and (ii) tell her a good example how the methodology can be applied to solve a public health problem.

Q1 Next page for re-worded Q1

1 Suppose X has a Binomial distribution with parameter n and p. Find the posterior density of p if we have a Beta(a,b) prior distribution for p. Calculate the mean and variance of the posterior density and describe their relationships with those of the Beta(a,b). Does it make sense to use method of moment estimators to estimate the mean and variance of this posterior density? Explain.

Sol. $X \sim Bin(n, p)$; $p \sim Beta(a, b)$.

(i) posterior of p:

$$f(p|x) \propto f(x|p)\, g(p)$$
$$= \left[\binom{n}{x} p^x (1-p)^{n-x}\right]\left[\frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}\right]$$
$$\propto p^{x+a-1}(1-p)^{n+b-x-1}$$

$\propto$ density of a $Beta(x+a, n+b-x)$ random variable.

$$\Rightarrow f(p|x) = \frac{1}{B(x+a, n+b-x)} p^{x+a-1}(1-p)^{n+b-x-1}$$

(ii) $$E(p|x) = \int_0^1 p f(p|x)\, dp$$
$$= \int_0^1 p^{x+a+1-1}(1-p)^{n+b-x-1} \frac{1}{B(x+a, n+b-x)}\, dp$$
$$= \frac{B(x+a+1, n+b-x)}{B(x+a, n+b-x)} = \frac{x+a}{n+a+b} \quad (a)$$

Similarly,
$$E(p^2|x) = \frac{(x+a)(x+a+1)}{(n+a+b+1)(n+a+b)}$$

Thus,
$$Var(p|x) = E(p^2|x) - (E(p|x))^2$$
$$= \frac{(n+b-x)(x+a)}{(n+a+b+1)(n+a+b)^2} \quad (b)$$

(iii): MOM:

$$\widehat{E(p|x)} = \int x\, dF_n(x)/n$$
$$= \frac{x}{n} \quad (\triangle)$$

where $F_n(\cdot)$ is the empirical dist. of X, and in this case,
$$F_n(t) = I(t \geq x).$$

Similarly,
$$\widehat{Var(p|x)} = \frac{1}{n}\frac{x}{n}\left(1-\frac{x}{n}\right)$$
$$= \frac{x(n-x)}{n^3} \quad (*)$$

Note that
$(\triangle)$ & (a), (b) & $(*)$
are equivalent in large sample sense.
Thus, it is reasonable to use $(\triangle)$ & $(*)$.

Q1 Reworded.

1 Suppose X|p has the Binomial distribution with parameter n and p and p has a Beta(a,b) density.

(i) Find the mean and variance of p.

(ii) Find the posterior of p|X and identify the density.

(iii) The predictive distribution of X is obtained by first writing down the joint density of X and p and integrating out p. Show that the predictive distribution of X is

$$P(X=x) = {}^{n}C_{x}\, B(a+x,b+n-x)/B(a,b),\ x=0,1,2,\dots,n.$$

This is the Beta-Binomial distribution with parameter n, a and b and is denoted by X~BetaBin(n,a,b).

(iv) Use your result in (i) and compute the mean and variance of the Beta-Binomial random variable with parameters n,a and b.

(v) Let s=a/(a+b) and show that results in (iv) can be, respectively, expressed as ns and ns(1-s)(a+b+n)/(a+b+1), thus showing that the variance of the Beta-Binomial variate has variance larger than the corresponding Binomial variate and they have the same mean.

Sol. (i): $f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$

$$\mathbb{E}p = \int_0^1 p f(p)\,dp = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a+1-1}(1-p)^{b-1}\,dp$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{a}{a+b}$$

$$\mathbb{E}p^2 = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} = \frac{a(a+1)}{(a+b)(a+b+1)}$$

$$\Rightarrow \mathrm{Var}\,p = \mathbb{E}p^2 - (\mathbb{E}p)^2 = \frac{ab}{(a+b)^2(a+b+1)}$$

(ii) $f(p|x) \propto f(x|p) f(p)$

$$\propto p^{x+a-1}(1-p)^{n+b-x-1}$$

$$\Rightarrow \mathbb{E}(p|x) = \frac{x+a}{n+a+b} \quad \&$$

$$f(p|x) = \frac{\Gamma(n+a+b+1)}{\Gamma(x+a+1)\Gamma(n-x+b)} p^{x+a-1}(1-p)^{n+b+x-1}$$

(iii): $f(x) = \int f(x|p) f(p)\,dp$

$$= \int \binom{n}{x} p^x (1-p)^{n-x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}\,dp$$

$$= \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a+x-1}(1-p)^{n+b-x-1}\,dp$$

$$= \binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+x)\Gamma(n+b-x)}{\Gamma(n+a+b-1)}$$

$$= \binom{n}{x} B(a+x, n+b-x)/B(a,b)$$

(iv): Let $X \sim \mathrm{BetaBin}(n,a,b)$

Then

$$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|p))$$

$$= \mathbb{E}(np)$$

$$= \frac{na}{a+b}$$

$$\mathrm{Var}\,X = \mathbb{E}(\mathrm{Var}(X|p)) + \mathrm{Var}(\mathbb{E}(X|p))$$

$$= \mathbb{E}(np - np^2) + \mathrm{Var}(np)$$

$$= \frac{na}{a+b} - \frac{na(a+1)}{(a+b)(a+b+1)} + \frac{n^2ab}{(a+b)^2(a+b+1)}$$

$$= \frac{nab(a+b+n)}{(a+b)^2(a+b+1)}$$

(v) $\mathbb{E}X = ns$. $\quad 1-s = \frac{b}{a+b}$

$$\mathrm{Var}\,X = \frac{ns(1-s)(a+b+n)}{a+b+1}$$

Since $\frac{a+b+n}{a+b+1} > 1$.

The variance is larger.

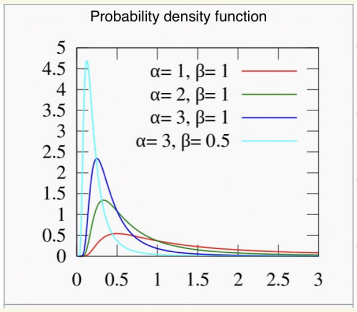
Q2:

2 Suppose X has an inverted gamma with density given by $f(x,a,b) = b^{a} x^{-a-1} \exp(-b/x)/\Gamma(a)$, where a>0 and b>0 and Γ is the Gamma function.

(i) Describe how values of a and b affect the density of X in terms of shape, skewness and symmetry.

(ii) Find its mean and describe how this density is typically used in estimating parameters in a linear model using a Bayesian paradigm.

Sol. (i)



skewness: $\frac{4\sqrt{\alpha-2}}{\alpha-3}$ for $\alpha>3$

kurtosis: $\frac{30\alpha-66}{(\alpha-3)(\alpha-4)}$ for $\alpha>4$

Thus, b has nothing to do with skewness & kurtosis.

(ii): $\mathbb{E}X = \int_0^\infty x f(x|a,b)\,dx$

$$= \int_0^\infty b^a x^{-a+1-1} e^{-\frac{b}{x}} \frac{1}{\Gamma(a)}\,dx$$

$$= \frac{b^a}{\Gamma(a)} \frac{\Gamma(a-1)}{b^{a-1}}$$

$$= \frac{b}{a-1} \text{ for } a>1.$$

LM: $Y = X\beta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$.

We would like to put a prior on (β, σ^2).

$$f(\beta, \sigma^2) = f(\beta|\sigma^2) f(\sigma^2)$$

& $f(\sigma^2)$ is an inverted gamma density, & $\beta|\sigma^2 \sim \mathcal{N}(m, \sigma^2 V)$.

Q3:

3 Suppose Z has a density proportional to $z^{n/2-1} \exp(-nz/2)$ and conditional on Z, X|Z=z is multivariate normal with mean 0 and covariance (1/z)I, where I is the k x k identity matrix. What is the density of X? Determine its mean and covariance matrix of X.

Sol. $X|Z=z \sim \mathcal{N}(0, \frac{1}{z}I)$

$$f_Z(z) = \frac{(\frac{n}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} z^{n/2-1} e^{-\frac{n}{2}z}$$

(i) density of X:

$$f_X(x) = \int_0^\infty f_{X|Z}(x) f_Z(z)\,dz$$

$$= \int_0^\infty (2\pi)^{-\frac{n}{2}} z^{\frac{n}{2}} \exp\left\{-\frac{z}{2}x^Tx\right\} \frac{(\frac{n}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} z^{\frac{n}{2}-1} e^{-\frac{n}{2}z}\,dz$$

$$\propto \int_0^\infty z^{n-1} e^{-\frac{1}{2}(x^Tx+n)z}\,dz$$

$$\propto \left(1+\frac{1}{n}x^Tx\right)^{-n} \quad (\triangle)$$

$(\triangle)$ is proportional to a multivariate t-dist.

$$t_n(0, I)$$

with $df=n$, $\mu=0$, $V=I$.

Thus, $X \sim t_n(0, I)$, density is

$$f_X(x) = \frac{\Gamma(n)}{(\pi n)^{\frac{n}{2}}\Gamma(\frac{n}{2})} \left(\left(1+\frac{1}{n}x^Tx\right)^{-\frac{n+n}{2}}\right.$$

(ii): mean & variance

Mean is trivial: $g(x) = x f_X(x)$ is an odd function, thus if exists, $\int g(x)\,dx = 0 \Rightarrow$ $\mathbb{E}X = 0$, provided $n \geq 2$.

Another method:

$$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|Z)) = \mathbb{E}(0) = 0.$$

For variance, see next page:

Let $X = (X_1, \dots, X_n)$.

$$\mathrm{Cov}(X_i, X_j) = \underbrace{\mathbb{E}(\mathrm{Cov}(X_i, X_j \mid Z))}_{①} - \underbrace{\mathrm{Cov}(\mathbb{E}(X_i \mid Z), \mathbb{E}(X_j \mid Z))}_{②}.$$

$② = \mathrm{Cov}(0, 0) = 0.$

$$① = \begin{cases} \mathbb{E}(0) & \text{if } i \neq j \\ \mathbb{E}\left(\frac{1}{Z}\right) & \text{if } i = j \end{cases}$$

Thus we have,

$$\mathrm{Var}\, X = \mathrm{Var}(\mathbb{E}(X \mid Z)) + \mathbb{E}(\mathrm{Var}(X \mid Z))$$

$$= I\, \mathbb{E}\left(\frac{1}{Z}\right).$$

$$\mathbb{E}\frac{1}{Z} = \int_0^\infty \frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} z^{\frac{n}{2}-1} \cdot e^{-\frac{n}{2}z} \cdot z^{-1}\, dz$$

$$= \frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \frac{\Gamma\left(\frac{n}{2}-1\right)}{\left(\frac{n}{2}\right)^{\frac{n}{2}-1}}$$

$$= \frac{\frac{n}{2}}{\frac{n}{2}-1} = \frac{n}{n-2}$$

$$\Rightarrow \mathrm{Var}\, X = \frac{n}{n-2} I$$

Q4:

2. Using the noninformative prior for $\boldsymbol{\theta}$, show that the conditional posterior density $f(\boldsymbol{\beta} \mid \mathbf{y}, \sigma)$ is multivariate normal. Hence deduce that the posterior mean of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}$.

Sol. $Y \mid \beta, \sigma^2 \sim \mathcal{N}_n(X\beta, \sigma^2 I)$.

density of $\theta^T = (\beta^T, \sigma)$: $f(\theta) \propto \frac{1}{\sigma}$

Then

$$f(\beta \mid Y, \sigma) \propto f(Y \mid \beta, \sigma) f(\theta)$$

$$\propto e^{-\frac{1}{2\sigma^2}(Y - X\beta)^T (Y - X\beta)}$$

However, let $X\hat{\beta} = P_X Y$, then

$$\|Y - X\beta\|_2^2 = \|Y - P_X Y + P_X Y - X\beta\|_2^2$$

$$= \|Y - P_X Y\|_2^2 + \|X(\hat{\beta} - \beta)\|_2^2 + 2\langle Y - P_X Y, P_X Y - X\beta \rangle$$

$$\overset{(\odot)}{=} \|Y - P_X Y\|_2^2 + \|X(\hat{\beta} - \beta)\|_2^2$$

$(\odot)$: $Y - P_X Y = (I - P_X) Y \in \mathrm{span}(X)^\perp$

$P_X Y - X\beta = X(\hat{\beta} - \beta) \in \mathrm{span}(X)$

$\Rightarrow \langle Y - P_X Y, P_X Y - X\beta \rangle = 0.$

$$\Rightarrow f(\beta \mid Y, \sigma) \propto e^{-\frac{1}{2\sigma^2}(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)}$$

Thus, $\beta \mid Y, \sigma \sim \mathcal{N}_p(\hat{\beta}, \sigma^2 (X^T X)^{-1})$.

$$\mathbb{E}(\beta \mid Y, \sigma) \overset{a.s.}{=} \hat{\beta}.$$

Q5:

3. Suppose that we use the noninformative prior for $\boldsymbol{\theta}$.

   (a) If $v = \sigma^2$, show that $f(v) \propto 1/v$.

   (b) Obtain an expression proportional to $f(\boldsymbol{\beta}, v|\mathbf{y})$.

   (c) Using (3.62), integrate out $\boldsymbol{\beta}$ to obtain

$$f(v|\mathbf{y}) \propto v^{-(\nu/2+1)} \exp\left(-\frac{a}{v}\right),$$

   where $\nu = n - p$ and $a = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2/2$.

   (d) Find the posterior mean of v.

Sol.

(a): $v = \sigma^2 \Rightarrow \log\sigma = \frac{1}{2}\log v$

But $f(\beta, \log\sigma) \propto 1$

Jacobian for $\begin{pmatrix}\beta \\ \log\sigma\end{pmatrix} \to \begin{pmatrix}\beta \\ v\end{pmatrix}$ is

$$J = \begin{bmatrix} I & \\ & \frac{1}{2v} \end{bmatrix}$$

Thus, density for $\begin{pmatrix}\beta \\ v\end{pmatrix}$ is

$$f(\beta, v) = |J| f(\beta, \log\sigma)$$
$$= \frac{1}{2v} \propto \frac{1}{v}$$

(b): $f(\beta, v|y)$

$$\propto f(y|\beta, v) f(\beta, v)$$
$$\propto f(y|\beta, v)\frac{1}{v}$$
$$= \frac{1}{v}\frac{1}{\sqrt{(2\pi v)^n}} e^{-\frac{1}{2v}(Y-X\beta)^T(Y-X\beta)}$$
$$\propto v^{-\frac{n}{2}-1} e^{-\frac{||Y-X\beta||_2^2}{2}(v^{-1})}$$

(c): Note we have

$$||Y-X\beta||_2^2 = ||Y-X\hat{\beta}||_2^2 + ||X(\hat{\beta}-\beta)||_2^2$$

Thus,

$$\int_{\mathbb{R}^p} e^{-\frac{1}{2v}||Y-X\beta||_2^2}\, d\beta$$
$$= \int_{\mathbb{R}^p} e^{-\frac{1}{2v}||Y-X\hat{\beta}||_2^2} e^{-\frac{1}{2v}(\beta-\hat{\beta})^T X^T X(\beta-\hat{\beta})}\, d\beta$$
$$= \frac{(2\pi)^{\frac{p}{2}} v^{\frac{p}{2}}}{\det|X^T X|} e^{-\frac{1}{2v}||Y-X\hat{\beta}||_2^2}$$

$\Rightarrow$

$$f(v|Y) = \int_{\mathbb{R}^p} f(\beta, v|y)\, d\beta$$
$$\propto v^{-\frac{n-p}{2}-1} e^{-\frac{1}{2v}||Y-X\hat{\beta}||_2^2}$$
$$\propto v^{-(\frac{\nu}{2}+1)} e^{-\frac{a}{v}}$$

(d): From (c), $v|Y \sim \mathrm{InvGa}(\frac{\nu}{2}, a)$

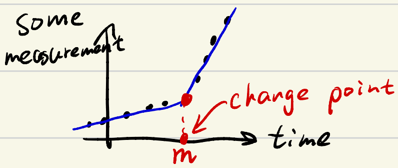
From Q2, we have

$$\mathbb{E}(v|Y) = \frac{a}{\frac{\nu}{2}-1}$$
$$= \frac{2a}{\nu-2}$$

Q6:

6 Last week, your supervisor came to you with the paper on Bayes Estimation of Two-Phase Linear Regression Model posted on Week 8 on the class website. She was very interested in the paper and want to apply the methodology in the paper to solve a public health problem. But she cannot understand much of the paper and have no idea how to apply the methodology to a public health problem. Your tasks are (i) explain to her in simple English in not more than ¾ page what the contents and methodology of the paper are about, and (ii) tell her a good example how the methodology can be applied to solve a public health problem.

(i): The paper discusss how to detect the change point (following figure) using Bayesian estimation.



The model is (m is the change point)

$$Y = \begin{cases} \alpha_1 + \beta_1 X_t + \varepsilon_1 & \text{if } t \le m \\ \alpha_2 + \beta_2 X_t + \varepsilon_2 & \text{else} \end{cases}$$

$$\varepsilon_1, \varepsilon_2 \overset{iid}{\sim} \mathcal{N}(0, 6^2)$$

Encoding prior information on $(\alpha_1, \alpha_2, \beta_1, \beta_2, 6^2)$ the authors calculate the posterior density of t:

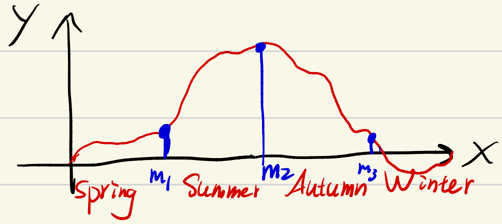
$$(\triangle) \quad f(t \mid Y) = \int_{\Theta} f(t, \alpha_1, \alpha_2, \beta_1, \beta_2, 6^2 \mid Y)\, d\theta$$

where $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, 6^2)$

Equation (17) & (26) explictly write down what $(\triangle)$ is.

With $f(t|y)$ & different choices of Criterion (AKA loss function), one can derive the estimation of the change point m. See eq (31)(32)(36)(39)(41)

(ii) Assuming Y represents the temperature & X represents date.



The above figure shows a typical pattern in a region of north-hemisphere. m_1, m_2, m_3 can be 3 potential "change points" of temperature.

One can apply the estimation of m in the paper to derive different estimated values of m_1, m_2 and m_3.