THE UNIVERSITY OF MELBOURNE

Student number Michaelle 998211

Semester 2 Special Assessment, 2021

School of Mathematics and Statistics

MAST30027 Modern Applied Statistics

Reading time: 30 minutes — Writing time: 2 hours — Upload time: 30 minutes

This exam consists of 12 pages (including this page) with 5 questions and 50 total marks

Page 1 of 12

add any extra pages after page 12

Page 1 of 12

Permitted Materials

- This exam and/or an offline electronic PDF reader, one or more copies of the masked exam template made available earlier, blank loose-leaf paper and a Casio FX-82 calculator.
- One double sided A4 page of notes (handwritten only).
- No headphones or earphones are permitted.

Instructions to Students

- Wave your hand right in front of your webcam if you wish to communicate with the supervisor at any time (before, during or after the exam).
- You must not be out of webcam view at any time without supervisor permission.
- You must not write your answers on an iPad or other electronic device.
- Off-line PDF readers (i) must have the screen visible in Zoom; (ii) must only be used to read exam questions (do not access other software or files); (iii) must be set in flight mode or have both internet and Bluetooth disabled as soon as the exam paper is downloaded.

Writing

- If you are writing answers on the exam or masked exam and need more space, use blank paper. Note this in the answer box, so the marker knows.
- If you are only writing on blank A4 paper, the first page must contain only your student number, subject code and subject name. Write on one side of each sheet only. Start each question on a new page and include the question number at the top of each page.

Scanning and Submitting

- **You must not leave Zoom supervision to scan your exam.** Put the pages in number order and the correct way up. Add any extra pages to the end. Use a scanning app to scan all pages to PDF. Scan directly from above. Crop pages to A4.
- Submit your scanned exam as a single PDF file and carefully review the submission in Gradescope. Scan again and resubmit if necessary. Do not leave Zoom supervision until you have confirmed orally with the supervisor that you have received the Gradescope confirmation email.
- **You must not submit or resubmit after having left Zoom supervision.**

Question 1 (10 marks)

(a) [redacted]

$f(x;\gamma) = \exp(\log(\gamma) - \gamma x)$

$= \exp(-\gamma x + \log(\gamma))$

by definition,

$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$

$= \exp\left(\frac{-\gamma x + \log(\gamma)}{a(\phi)} + 0\right)$

to prove that Exponential distribution is an exponential family

$\theta = -\gamma$

$b(\theta) = b(-\gamma) = \log(-1/-\gamma)$

$= \log(-1/\theta)$

$a(\phi) = \phi$

$\phi = 1$

$c(y,\phi) = 0$

Variance function,

$V(X) = V(\mu_i)\, a(\phi) = b''(\theta)\, a(\phi)$

$\frac{1}{\gamma^2} = b''(\theta)$, $V(\mu) = \frac{1/\gamma^2}{\phi} = 1/\gamma^2 = \left(\frac{1}{\gamma}\right)^2 = \mu^2$ final answer.

(b) [redacted]

$$P(f(x_i;\mu,\sigma^2)) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{(x_i-1)^2}{2\tau}\right)$$

$N(1,\tau)$
$\mu \quad \sigma^2$

likelihood $\rightarrow$
$$= \left(\frac{1}{2\pi\tau}\right)^{+n/2} \exp\left(-\frac{1}{2\tau}\sum_{i=1}^{n}(x_i-1)^2\right)$$

$$\log P(f(x_i;\mu,\sigma^2)) = \frac{n}{-2}\log(2\pi\tau) - \frac{1}{2\tau}\sum_{i=1}^{n}(x_i-1)^2$$

log likelihood, $\qquad \tau^{-1} \quad -\tau^{-2}$

$$\frac{\partial \log P(f(x_i;\mu,\sigma^2))}{\partial \lambda} = \frac{-n}{2}\frac{1}{\cancel{2\pi}\tau}\cancel{2\pi} - \frac{1}{2}\left(-\frac{1}{\tau^2}\right)\sum_{i=1}^{n}(x_i-1)^2 = 0$$

$$\Rightarrow \frac{-n}{2\tau} + \frac{1}{2\tau^2}\sum_{i=1}^{n}(x_i-1)^2 = 0$$

$$\Rightarrow \frac{n}{\cancel{2}\tau} = \frac{1}{\cancel{2}\tau^2}\sum_{i=1}^{n}(x_i-1)^2$$

τ^{1}
$-\tau^{-2}$

$$\hat{\tau} = \frac{\sum_{i=1}^{n}(x_i-1)^2}{n} \quad \text{MLE},$$

τ^{-2}
$-2\tau^{-3}$
τ^{1}
$-\tau^{-2}$

$$I(\lambda) = -E\left[\frac{\partial^2 \log[P(f(x_i;\mu,\sigma^2))]}{\partial \lambda^2}\right]$$

$$= -E\left[\frac{+n}{2\tau^2} + \frac{1}{2}(-2)\frac{1}{\tau^3}\sum_{i=1}^{n}(x_i-1)^2\right] \leftarrow \text{Fisher Information.}$$

$= \cancel{\frac{-n}{2\tau}}$ $\rightarrow$ MORE TO continue)

Q1(b)

$$I(\lambda) = -\mathbb{E}\left[\frac{\partial^2 \log[p(f(x_i;\mu,\sigma^2))]}{\partial \lambda^2}\right]$$

$$= -\mathbb{E}\left[\frac{n}{2\tau^2} + \frac{1}{2}(-2)\frac{1}{\tau^3}\sum_{i=1}^{n}(x_i-1)^2\right]$$

$$= -\frac{n}{2\tau^2} + \frac{1}{\tau^3}\mathbb{E}\left[\sum_{i=1}^{n}(x_i-1)^2\right]$$

$$= -\frac{n}{2\tau^2} + \frac{1}{\tau^3}\tau n = -\frac{n}{2\tau^2} + \frac{2n}{2\tau^2} = \frac{n}{2\tau^2}.$$

$$V(\lambda) = \frac{1}{I(\lambda)} = \frac{2\tau^2}{n}$$ (take the inverse of the Fisher Information)

Then $\hat{\lambda} \sim N\left(\lambda, \frac{2\tau^2}{n}\right)$ Asymptotic distribution.

Question 2 (4 marks)

* Variance is larger than ~~in~~ the response variable doctorco, ~~changes the deviance for the residuals~~ (i.e. Var(Y_i) is larger)

· Using $\phi = 1$, (ignoring overdispersion).
* leads to insignificant values for the parameters sex, age, income, freepoor. which leads the model to have interference, in case we don't have a likelihood, when the generalized linear model. (see bottom)

Ignoring over fitting when fitting a Poisson regression
~~Where deviances are different~~ ~~depending~~

Estimation is fixed, but your interference has changed, with the proper modelling F-statistic is reduced, making the model comparison less significant in general. Resulting fewer significant variables for sex, age, income, freepoor in the model. Changing variance can also change a wider Confidence Interval for your parameter estimates.

Question 3 (11 marks)

$z_i \sim \text{cate}(p, 1-p)$, $\boxed{\theta = (p, \tau)}$

$x_i | z_i = 1 \sim N(0, 1^2)$

$x_i | z_i = 2 \sim N(0, 1/\tau)$

First, $p(x_i, z_i | \theta) = \prod_{i=1}^{n} p(x_i | z_i, \theta)\, p(z_i|\theta)$

$= \prod_{i=1}^{n} \prod_{k=1}^{2} p(x_i | z_i = k, \theta)\, p(z_i = k | \theta)^{I(z_i = k)}$

$\log p(x_i, z_i | \theta) = \sum_{i=1}^{n} \sum_{k=1}^{2} I(z_i = k)$ ~~$p(x_i|z_i$~~

$[\log p(x_i | z_i = k, \theta) + \log p(z_i = k | \theta)]$

$Q(\theta, \theta^0) = E_{z|x,\theta^0}[\log p(x, z | \theta)]$

$= \sum_{i=1}^{n} \sum_{k=1}^{2} p(z_i = k | x_i, \theta^0) \Big[-\frac{n}{2}\log(2\pi) - \frac{1}{2}(x_i)^2$

$- \frac{1}{2}\log\left(\frac{2\pi}{\tau}\right) - \frac{\tau}{2}(x_i)^2$

$+ \log p$

$+ \log 1-p \Big]$

→ continue next page

$\frac{1}{\sqrt{2\pi 1}} \exp\left(-\frac{1}{2}(x_i - 0)^2\right) = -\frac{n}{2}\log(2\pi)$

$\frac{1}{\sqrt{2\pi \frac{1}{\tau}}} \exp\left(-\frac{\tau}{2}(x_i - 0)^2\right)$

There is extra answer space for Question 3 on the next page

Additional answer space only for Question 3—submit this page even if not used

$Q(\theta, \theta^0) = \sum_{i=1}^{n} \sum_{k=1}^{2} P(Z_i = k \mid X_i, \theta^0) \left[-\frac{n}{2}\log(2\pi) - \frac{1}{2}(x_i - \mu)^2 \right.$
$\left. \underbrace{-\frac{n}{2}\log\left(\frac{2\pi}{\tau}\right)}_{} - \frac{\tau}{2}(x_i - \mu)^2 + \log p + \log(1-p) \right]$.

$\log(u)$

$u = 1-p$ $\quad \frac{1}{u} \times -1$

$\boxed{u' = -1}$ $\quad = -1/u$

M-step,

$\frac{\partial Q(\theta, \theta^0)}{\partial p} = \sum_{i=1}^{n} \left[\frac{P(Z_i = 1 \mid X_i, \theta^0)}{p} - \frac{P(Z_i = 2 \mid X_i, \theta^0)}{1-p} \right]$
$= 0$,

$\boxed{\text{NOTE } P(Z_i = 2 \mid X_i, \theta^0) = 1 - P(Z_i = 1 \mid X_i, \theta^0)}$

$\Rightarrow \cancel{\sum_{i=1}^{n} P(Z_i = 1 \mid X_i, \theta^0)}$ $\cancel{n(1-p)}$

$\sum_{i=1}^{n} \frac{(1-p) P(Z_i = 1 \mid X_i, \theta^0) - p (1 - P(Z_i = 1 \mid X_i, \theta^0))}{p(1-p)} = 0$

$n(1-p) \sum_{i=1}^{n} P(Z_i = 1 \mid X_i, \theta^0) + np \sum_{i=1}^{n} P(Z_i = 1 \mid X_i, \theta^0) - np = 0$

$\cancel{n - np \sum_i}$

$n \sum_{i=1}^{n} P(Z_i = 1 \mid X_i, \theta^0) - np = 0$

$\to$ see separate page

$\hat{p} = \frac{\sum_{i=1}^{n} P(Z_i = 1 \mid X_i, \theta^0)}{n}$

LOOK AWAY

NOTE: $-\frac{n}{2}\log\left(\frac{2\pi}{\tau}\right) = -\frac{n}{2}(\log(2\pi) - \log(\tau))$
$= -n/2 \log(2\pi) + n/2 \log(\tau)$.

$$\sum_{i=1}^{n} (1-p)\,\mathbb{P}(z_i=1|x_i,\theta^0) - p\,(1-\mathbb{P}(z_i=1|x_i,\theta^0))$$

$$= \sum_{i=1}^{n} \Big(\mathbb{P}(z_i=1|x_i,\theta^0) - \cancel{p\,\mathbb{P}(z_i=1|x_i,\theta^0)}$$

$$- p + \cancel{p\,\mathbb{P}(z_i=1|x_i,\theta^0)}\Big)$$

$$= \sum_{i=1}^{n} \mathbb{P}(z_i=1|x_i,\theta^0) - np = 0$$

$$\Rightarrow \quad np = \sum_{i=1}^{n} \mathbb{P}(z_i=1|x_i,\theta^0)$$

$$\boxed{\hat{p} = \frac{\sum_{i=1}^{n} \mathbb{P}(z_i=1|x_i,\theta^0)}{n}}$$ — Final Answer

$$\frac{\partial Q(\theta,\theta^0)}{\partial \tau} = \sum_{i=1}^{n} \Big[\mathbb{P}(z_i=2|x_i,\theta^0)\,\big(\;(*)\;\big)\Big] = 0$$

$$(*) = \frac{n}{2\tau} - \frac{1}{2}x_i^2$$

$$a = \frac{2\pi}{\tau}$$

$$\sum_{i=1}^{n} \mathbb{P}(z_i=2|x_i,\theta^0)\left(\frac{n}{2\tau} - \frac{1}{2}x_i^2\right) = 0$$

$$\sum_{i=1}^{n} \mathbb{P}(z_i=2|x_i,\theta^0)\left(\frac{n-\tau x_i^2}{\cancel{2\tau}}\right) = 0$$ → Denominator disappears,

$$n\sum_{i=1}^{n} \mathbb{P}(z_i=2|x_i,\theta^0) = \tau\sum_{i=1}^{n} x_i^2\,\mathbb{P}(z_i=2|x_i,\theta^0)$$

$$\boxed{\hat{\tau} = \frac{n\sum_{i=1}^{n} \mathbb{P}(z_i=2|x_i,\theta^0)}{\sum_{i=1}^{n} x_i^2\,\mathbb{P}(z_i=2|x_i,\theta^0)}}$$ — Final Answer.

Page 8 of 12 — add any extra pages after page 12 — Page 8 of 12

Question 4 (12 marks)

$$p(\mu, r=1 \mid X=2) + p(\mu, r=0 \mid X=2)$$

$$= \cancel{\ldots}$$

$$\underbrace{= \frac{1}{\sqrt{2\pi 1^2}} \exp\left(-\frac{(\mu-1)^2}{2\times 1^2} - \frac{1}{2}\left(\frac{(\cancel{x}-\mu)^2}{2\times 1^2}\right)\right) + \frac{1}{\sqrt{2\pi 1^2}} \exp\left(\cancel{-\frac{\mu}{2\times 1^2}} - \frac{1}{2}\left(\frac{(x-\mu)^2}{2}\right)\right)}_{\text{linear combination of normal distributions.}}$$

There is extra answer space for Question 4 on the next page

Additional answer space only for Question 4—submit this page even if not used

① start w/ an initial value of $\mu^{(0)}$ set $t=0$,

② sample $\mu^{(t+1)}$ from $p(\mu \mid X=2)$ using the full condition distribution as the proposal distribution

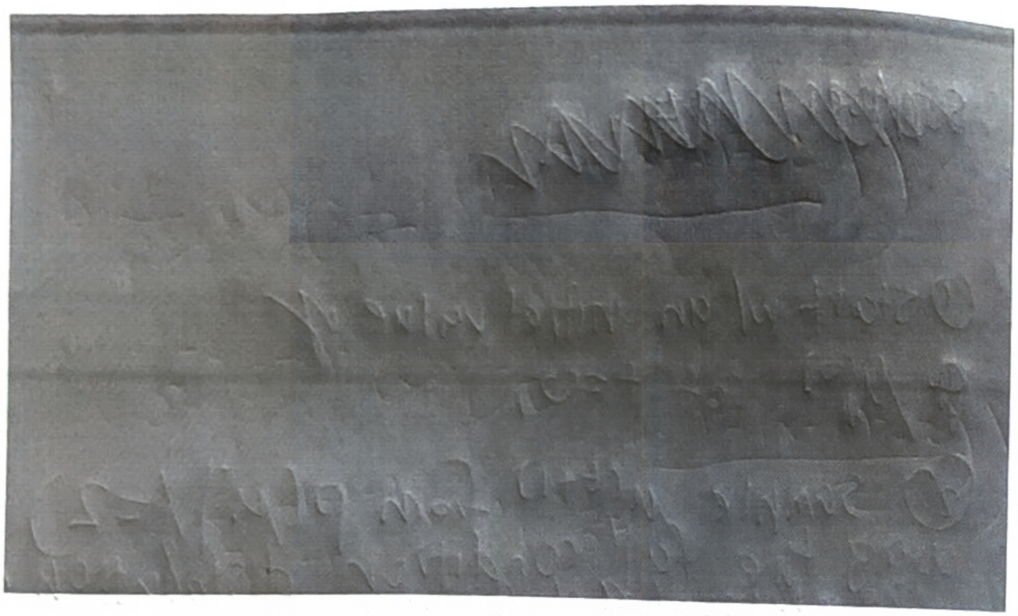
③ same $\mu_2^{(t+1)}$ from $p(\mu_2 \mid X, \mu_1^{(t+1)}, \ldots, \mu_k^{(t)})$,

④ ... ⑤ sample $\mu_k^{(t+1)}$ from $p(\mu_k \mid X=2)$, gives an acceptance prob. of $A=1$,

~~→ will slow down the Gibbs sampler~~

Question 5 (13 marks)

(a)



$(\mu_c, \sigma_c^2) = (-1, 2^2)$, $(\mu_n, \sigma_n^2) = (1, 1^2)$

$x = 0$,

$$AP = \min\left\{1, \frac{p(\mu_n, \sigma_n^2, x)^{(1)}}{p(\mu_c, \sigma_c^2, x)_{(2)}} \frac{q(\mu_n, \sigma_n^2 | \mu_c, \sigma_c^2)^{(3)}}{q(\mu_c, \sigma_c^2 | \mu_n, \sigma_n^2)_{(4)}}\right\}$$

$$p(\mu, \sigma^2; x) \propto \frac{1}{\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$\mu_n, \sigma_n^2 = (1, 1^2)$, $x = 0$

$$= \frac{1}{1^2} \frac{1}{\sqrt{2\pi 1}} \exp\left(-\frac{(0-1)^2}{2(1)^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(1)^2}{2}\right)$$

$\mu_c, \sigma_c^2 = (-1, 2^2)$, then

$$\frac{1}{\sqrt{2\pi 2^2}} \exp\left(-\frac{(+1)^2}{2 \cdot 2^2}\right)$$

$$=\frac{(1)}{(2)}=\frac{\cancel{\frac{1}{\sqrt{2\pi}}}}{\cancel{\frac{1}{2\sqrt{2\pi}}}}\exp\left(-\frac{1}{2}+\frac{1}{8}\right)$$

$$=\frac{1}{2}\exp\left(-\frac{3}{8}\right).$$

$$(3)=q(\mu_n,\sigma_n^2,\mu_c,\sigma_c^2)=\underbrace{q(\mu_n,\mu_c)}_{(3a)}\,\underbrace{q(\sigma_n^2,\sigma_c^2)}_{(3b)}$$

$$(4)=q(\mu_c,\sigma_c^2,\mu_n,\sigma_n^2)=\underbrace{q(\mu_c,\mu_n)}_{(4a)}\,\underbrace{q(\sigma_c^2,\sigma_n^2)}_{(4b)}.$$

$$(3a)=q(\mu_n,\mu_c)=\cancel{\frac{1}{\sqrt{2\pi\sigma_n^2}}}\exp\left(-\frac{(\mu_c-\mu_c)^2}{2\sigma_n^2}\right)$$

$$(4a)=q(\mu_c,\mu_n)=\cancel{\frac{1}{\sqrt{2\pi\sigma_n^2}}}\exp\left(-\frac{(\mu_n-\mu_c)^2}{2\sigma_n^2}\right)$$

$$\frac{(3a)}{(4a)}=\cancel{\frac{1}{\sqrt{2\pi}}}\exp\left(0+\frac{(1-(-1))^2}{2(1)^2}\right)$$

$$=\exp\left(\frac{4}{2}\right)=\exp(2).$$

$$\frac{(3b)}{(4b)}=\frac{\cancel{\frac{1}{\sigma_c^2}}\exp\left(-\frac{\sigma_c^2}{\sigma_c^2}\right)}{\cancel{\frac{1}{\sigma_c^2}}\exp\left(-\frac{\sigma_n^2}{\sigma_c^2}\right)}$$

$$=\exp\left(-1+\frac{1^2}{2^2}\right)=\exp\left(-1+\frac{1}{4}\right)=\exp\left(-\frac{3}{4}\right)$$

then, $\frac{(1)}{(2)} \frac{(3a)}{(4a)} \frac{(3b)}{(4b)}$

$$= \frac{1}{2}\exp\left(-\frac{3}{8}\right)\frac{\exp(2)}{\exp(-3/4)}$$

$$= \frac{1}{2}\exp\left(\left(-\frac{3}{8}\right)+\left(\frac{6}{8}\right)+2\right)$$

$$AP = \min\left(1, \frac{1}{2}\exp\left(\frac{19}{2}\right)\right) = \underline{\underline{1}}$$

add any extra pages after page 12 -- Page 11 of 12

(b)

$$p(\mu, \sigma^2, x) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mu-1)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$(\sigma^2)^{-1-1} \exp\left(-\frac{1}{\sigma^2}\right)$$

$$\propto \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}\left((\mu-1)^2 + (x-\mu)^2 + 2\right)\right)$$

$$(\sigma^2)^{-1-1}$$

$$= \cancel{\frac{1}{2\pi}} \propto \frac{\sigma^{-2}}{(\sigma^2)^2} \exp\left(-\frac{1}{\sigma^2}\left(\frac{(\mu-1)^2 + (x-\mu)^2 + 2}{2}\right)\right)$$

continuing next page

There is extra answer space for Question 5 on the next page

Additional answer space only for Question 5—submit this page even if not used

$$\propto \frac{1}{\sigma^2 \sigma^4} \exp\left(-\frac{1}{\sigma^2}\left(\frac{(\mu-1)^2+(x-\mu)^2+2}{2}\right)\right)$$

$$\propto (\sigma^2)^{-1-1-1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{(\mu-1)^2+(x-\mu)^2+2}{2}\right)\right)$$

$$= (\sigma^2)^{-2-1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{(\mu-1)^2+(x-\mu)^2+2}{2}\right)\right).$$

$$\log q_\mu^*(\mu) \propto E_{\sigma^2}\left[(-2-1)\log(\sigma^2) - \frac{1}{\sigma^2}\left(\frac{(\mu-1)^2+(x-\mu)^2+2}{2}\right)\right]$$

$$\propto (-3)\, E_{\sigma^2}\log(\sigma^2) - \frac{1}{E_{\sigma^2}(\sigma^2)}\left[\frac{(\mu-1)^2+(x-\mu)^2+2}{2}\right]$$

$$q_\mu^*(\mu) \propto \underbrace{E_{\sigma^2}[\sigma^2]^{-2-1}}_{\propto E_{\sigma^2}[\mu^2]} \exp\left(-\frac{1}{E_{\sigma^2}(\sigma^2)}\left[\frac{(\mu-1)^2+(x-\mu)^2+2}{2}\right]\right)$$

$q_\mu^*(\mu)$ pdf is

Inverse Gamma$\left(2, \frac{(\mu-1)^2+(x-\mu)^2+2}{2}\right)$

observe x is 0,

$\to \left(2, \frac{(\mu^2-2\mu+1+\mu^2+2)}{2}\right)$

similarly,

$q_{\sigma^2}^*(\sigma^2)$ pdf is

Inverse Gamma$\left(2, \frac{E_\mu[(\mu-1)^2]+E_\mu[(x-\mu)^2]+2}{2}\right)$ (since $x=0$)

What we care

$\approx$ Inverse Gamma$\left(2, \frac{E_\mu[(\mu-1)^2+E_\mu(\mu^2)+2}{2}\right)$

$\approx$ Inverse Gamma $\left(2, \frac{2\mu^2-2\mu+3}{2}\right)$

End of Exam — Total Available Marks = 50

$\approx$ Inverse Gamma$\left(2, \left(E_\mu(\mu^2)-2E_\mu(\mu)+3+E_\mu(\mu^2)\right)/2\right)$

$\boxed{V_\mu(\mu) = E_\mu(\mu^2) - E_\mu(\mu)^2}$

$E_\mu(\mu^2) = E$

$\propto$ Inverse Gamma$\left(2, \frac{2E_\mu(\mu^2)-2E_\mu(\mu)+3}{2}\right)$

$\approx$ Inverse Gamma (2, Next page).

continuing Q5 b)

$$V_\mu(\mu) = \mathbb{E}_\mu(\mu^2) - (\mathbb{E}_\mu(\mu))^2$$

$$\mathbb{E}_\mu(\mu^2) = V_\mu(\mu) + \mathbb{E}_\mu(\mu)^2$$

then,

$q^*_{\sigma^2}(\sigma^2)$ pdf is

Inverse Gamma $\left(2, \frac{1}{2}\left(2\left(V_\mu(\mu) + \mathbb{E}_\mu(\mu)^2\right) - 2\mathbb{E}_\mu(\mu) + 3\right)\right)$

Final answer,

$q^*_\mu(\mu)$ pdf is Inverse Gamma$\left(2, \frac{2\mu^2 - 2\mu + 3}{2}\right)$

$q^*_{\sigma^2}(\sigma^2)$ pdf is Inverse Gamma$\left(2, \frac{1}{2}\left(2V_\mu(\mu) + 2(\mathbb{E}_\mu(\mu))^2 - 2\mathbb{E}_\mu(\mu) + 3\right)\right)$