# FALL 2016

# HUNTER COLLEGE

# STAT 706

# General Linear Model I

# Final Exam

Last Name: ______________________________

First Name: ______________________________

December 21, 2016

Instructions

- There are 4 questions, each question is between 5-7 points. The maximal score is 24 points. A perfect score is **20** points.
- There are two versions of Question 2, 3, 4 with different points. Please choice one version and only one to answer.
- Show all work. You will receive partial credit for partially completed problems.
- You may use any references.

**Q1**. (5pt) An exponentially tilted or exponential twisting (ET) density $f_{\mathrm{ET}}$ based on a base density $f_0$ and with natural parameter $\theta$ is given by $f_{\mathrm{ET}} = \exp(z\theta)f_0(z)/M(\theta)$, where $M(\theta)$ is the moment generating function (MGF) of the base density $f_0$. Take $f_0$ to be the hyperbolic secant density $f_0(z) = \frac{1}{2}\mathtt{sech}(\pi z/2)$, $z \in (-\infty, +\infty)$, with the the moment generating function $M(\theta) = \mathtt{sec}(\theta)$, $\theta \in (-\pi/2, +\pi/2)$

(1) Verify that the distribution $f_{ET}$ generated by $f_0$ belongs to an exponential family.

(2) Find the mean and variance of this distribution. Express the variance as a function of the mean.

(3) Find the canonical link for a generalized linear model (GLM) from the exponentially tilted distribution $f_{ET}$.

**Q2**. Please **choose only one** question from 2-A and 2-B. Circle your option.

**2-A**. (5 pts) Suppose that $\{Z_i\}_{i=1}^n$ are independent with $\mu_i = \mathbb{E}(Z_i)$ satisfying $\log \mu_i = x_i\beta$, where $x_i$ is the univariate predictor, and with $\texttt{var}(Z_i) = \mu/2$.

(1) Give the quasi-likelihood estimating equation for $\beta$ and find the asymptotic variance of $\widehat{\beta}_{MQL}$, the maximum quasi-likelihood estimator (MQLE) of $\beta$.

(2) Assuming that $\{Z_i\}_{i=1}^n$ are normally distributed (with means and variances as given above). Give the likelihood estimating equation for $\beta$ and find the asymptotic variance of $\widehat{\beta}_{ML}$, the maximum likelihood estimator (MLE) of $\beta$.

(3) Calculate the asymptotic relative efficiency of $\widehat{\beta}_{ML}$ with respect to $\widehat{\beta}_{MQL}$, which is defined as $\widehat{\sigma}^2_{\widehat{\beta}_{MQL}}/\widehat{\sigma}^2_{\widehat{\beta}_{ML}}$. Is the asymptotic relative efficiency greater or less than 1? Explain it.

**2-B**. (6 pts) Suppose that $\{Z_i\}_{i=1}^n$ are independent with $\mu_i = \mathbb{E}(Z_i)$ satisfying $\log \mu_i = x_i\beta$, where $x_i$ is the univariate predictor, and with $\texttt{var}(Z_i) = \phi\mu^2$, where $\phi > 0$ is a constant.

(1) Give the quasi-likelihood estimating equation for $\beta$ and find the asymptotic variance of $\widehat{\beta}_{MQL}$, the maximum quasi-likelihood estimator (MQLE) of $\beta$.

(2) Assuming that $\{Z_i\}_{i=1}^n$ are normally distributed (with means and variances as given above). Give the likelihood estimating equation for $\beta$ and find the asymptotic variance of $\widehat{\beta}_{ML}$, the maximum likelihood estimator (MLE) of $\beta$.

**Q3**. Please **choose only one** question from 3-A and 3-B. Circle your option.
The Kalman filter is a recursive estimator used to estimate the state of a linear time-varying state equation, in which the states are driven by noise and observations are made in the presence of noise. By using the Kalman filter, a sequence of estimates of the state of the system $x_t$ is determined by giving a sequence of measurements $\{y_0, y_1, \cdots\}$.

Consider a univariate discrete-time state-variable system driven by noise, with noisy observations

$$
\begin{aligned}
x_{t+1} &= a_t x_t + w_t, \\
y_t &= c_t x_t + v_t,
\end{aligned}
$$

where the state noise $w_t \sim N(0, \sigma_t^2)$, and the observation noise $v_t \sim N(0, \tau_t^2)$. Noises $\{w_0, w_1, \cdots, v_0, v_1, \cdots\}$ are mutually independent.

**3-A**. (5 pts) Write down the steps to update the estimate of the state.
Verify that $P_{t+1|t+1} > 0$ if $P_{t|t} > 0$.
Suppose that $a_t = 3$, $c_t = 1$, $\sigma_t^2 = 1^2$, $\tau_t^2 = 2^2$. Based on two observations at $t = 0$ and $t = 1$, which are $y_0 = 1$, $y_1 = 2$, starting from an initial estimate $\widehat{x}_{0|-1} = 0$, with an initial variance $P_{-1|-1} = 1^2$, use the Kalman filter to calculate $\widehat{x}_{2|1}$, which is the estimate of the state at $t = 2$ by giving the observations $\mathcal{Y}_1 = \{y_0, y_1\}$.

**3-B**. (7 pts) Do option A at first. Keep $a_t$, $c_t$, $y_0$, $y_1$, $\widehat{x}_{0|-1}$, $P_{-1|-1}$ the same, and consider three more cases with (1) $\sigma_t^2 = 1^2$, $\tau_t^2 = 1^2$, (2) $\sigma_t^2 = 2^2$, $\tau_t^2 = 1^2$, and (3) $\sigma_t^2 = 2^2$, $\tau_t^2 = 2^2$. Including the case in option A, there are totally four cases. Compare the estimate $\widehat{x}_{2|1}$ and $P_{2|1}$ in all four cases. Discuss the difference.

**Q4**. Please **choose only one** question from 4-A and 4-B. Circle your option.
**4-A**. (5 pts) Researchers record how long the flu last for patients in two groups.

| Group | Days until recovery |
|---|---|
| Control | 5, 9 |
| Treatment | 2, 4, 6 |

Use one variable, with takes values of 0 and 1, being an indicator variable for the two sample. Then the hazard in samples 0 and 1 are respectively $\lambda_0(t)$ and $e^{\beta}\lambda_0(t)$.

Use the Cox regression model to (1) find the estimating equation of $\beta$, and (2) test the null hypothesis $H_0 : \beta = 0$, report the $p$-value and draw a conclusion.
**4-B**. (6 pts) Do option A at first.
Do a two-sample $t$-test on this data set, and report the $p$-value. Show how to calculate the $t$-statistic, the degree of freedom, and the $p$-value, please do not use software to directly get the final results. Compare two results by using the Cox model and the $t$-test. Discuss the difference.

More space

More space

End of the final exam of Stat 706 (Instructor: Jiangtao Gou)