Coursework 3: STAT 570

Philip Pham

October 15, 2018

1. Consider the Poisson-gamma random effects model given by

$$Y_i \mid \mu_i, \theta_i \sim \text{Poisson}\left(\mu_i\theta_i\right) \tag{1}$$
$$\theta_i \sim \text{Gamma}\left(b, b\right), \tag{2}$$

which leads to a negative binomial marginal model with the variance a quadratic function of the mean. Design a simulation study, along the lines of that which produced Table 2.3 in the book (overdispersed Poisson example) to investigate the efficiency and robustness under

- a Poisson model;
- quasi-likelihood with $\mathbb{E}\left[Y\right] = \mu$ and $\text{Var}\left(Y\right) = \alpha\mu$; and
- sandwich estimation.

Use a log-linear model

$$\log \mu_i = \beta_0 + \beta_1 x_i, \tag{3}$$

with $x_i \sim_{\text{iid}} \mathcal{N}\left(0, 1\right)$ for $i = 1, 2, \ldots, n$, and $\beta_0 = -2$ and $\beta_1 = \log 2$.

Simulate for:

- $b \in \{0.2, 1, 10, 1000\}$.
- $n \in \{10, 20, 50, 100, 250\}$.

Summarize what your take away message is after carrying out these simulations.

Solution: Note that

$$\begin{aligned}
\mathbb{P}\left(Y_i = y \mid \mu_i\right) &= \int_0^\infty \mathbb{P}\left(Y_i = y \mid \mu_i, \theta_i = \theta\right)\mathbb{P}\left(\theta_i = \theta \mid b\right)\,\mathrm{d}\theta \\
&= \int_0^\infty \left(\frac{(\mu_i\theta)^y}{y!}\exp\left(-\mu_i\theta\right)\right)\left(\frac{b^b}{\Gamma(b)}\theta^{b-1}\exp\left(-b\theta\right)\right)\,\mathrm{d}\theta \\
&= \frac{\mu_i^y b^b}{y!\Gamma(b)}\int_0^\infty \theta^{b+y-1}\exp\left(-\theta(b+\mu_i)\right)\,\mathrm{d}\theta \\
&= \frac{\Gamma(y+b)}{y!\Gamma(b)}\frac{\mu_i^y b^b}{(\mu_i+b)^{b+y}} = \frac{\Gamma(y+b)}{y!\Gamma(b)}\left(\frac{b}{\mu_i+b}\right)^b\left(\frac{\mu_i}{\mu_i+b}\right)^y \\
&\sim \text{NegativeBinomial}\left(b, \frac{\mu_i}{\mu_i+b}\right).
\end{aligned} \tag{4}$$

By properties of the negative binomial distribution, we have that

$$\begin{aligned}\mathbb{E}\left[Y_i \mid x_i\right] &= \mu_i = \exp\left(\beta_0 + \beta_1 x_i\right) \\ \operatorname{Var}\left(Y_i \mid x_i\right) &= \mu_i\left(1 + \frac{\mu_i}{b}\right).\end{aligned} \tag{5}$$

Thus, smaller values of b correspond to more dispersion.

Poisson Model

In the Poisson model, we assume that $\operatorname{Var}\left(Y_i \mid x_i\right) = \mu_i$, e.g. $b \to \infty$, so we neglect the overdispersion parameter.
In this case, the log-likelihood function is

$$l\left(\beta\right) = \sum_{i=1}^{n}\left[y_i\left(\beta_0 + \beta_1 x_i\right) - \exp\left(\beta_0 + \beta_1 x_i\right) - \sum_{k=1}^{y_i}\log k\right], \tag{6}$$

which gives us the score function

$$S\left(\beta\right) = \sum_{i=1}^{n}\begin{pmatrix} y_i - \exp\left(\beta_0 + \beta_1 x_i\right) \\ x_i y_i - x_i \exp\left(\beta_0 + \beta_1 x_i\right)\end{pmatrix}. \tag{7}$$

We can estimate β by solving for $S\left(\hat{\beta}\right) = \mathbf{0}$, numerically.
We can estimate the variance of the estimates from the Fisher information,

$$\begin{aligned}\operatorname{Var}\left(\hat{\beta}\right) &\approx I_n\left(\hat{\beta}\right)^{-1} \\ &= \left(\sum_{i=1}^{n}\begin{pmatrix}\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) & x_i\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) \\ x_i\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) & x_i^2\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\end{pmatrix}\right)^{-1} \\ &= \frac{1}{\left(\sum_{i=1}^{n}\hat{\mu}_i\right)\left(\sum_{i=1}^{n}x_i^2\hat{\mu}_i\right) - \left(\sum_{i=1}^{n}x_i\hat{\mu}_i\right)^2}\begin{pmatrix}\sum_{i=1}^{n}x_i^2\hat{\mu}_i & -\sum_{i=1}^{n}x_i\hat{\mu}_i \\ -\sum_{i=1}^{n}x_i\hat{\mu}_i & \sum_{i=1}^{n}\hat{\mu}_i\end{pmatrix},\end{aligned} \tag{8}$$

where $\hat{\mu}_i = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$.

Quasi-likelihood

In a quasi-likelihood model, we specify the mean and variance as

$$\begin{aligned}\mathbb{E}\left[Y_i \mid x_i\right] &= \mu_i = \exp\left(\beta_0 + \beta_1 x_i\right) \\ \operatorname{Var}\left(Y_i \mid x_i\right) &= \alpha\mu_i\end{aligned} \tag{9}$$

From Equation 5, we see that this is not quite correct, still, but it is closer to the real model than the Poisson model.
Then, by Equation 2.30 of Wakefield's *Bayesian and Frequentist Regression Methods* our estimating function is

$$\begin{aligned}U\left(\beta\right) &= D^{\intercal}V^{-1}\left(y - \mu\right)/\alpha \\ &= \sum_{i=1}^{n}\begin{pmatrix}\exp\left(\beta_0 + \beta_1 x_i\right) \\ x_i\exp\left(\beta_0 + \beta_1 x_i\right)\end{pmatrix}\frac{y_i - \exp\left(\beta_0 + \beta_1 x_i\right)}{\alpha\exp\left(\beta_0 + \beta_1 x_i\right)} \\ &= \frac{1}{\alpha}\sum_{i=1}^{n}\begin{pmatrix} y_i - \exp\left(\beta_0 + \beta_1 x_i\right) \\ x_i y_i - x_i\exp\left(\beta_0 + \beta_1 x_i\right)\end{pmatrix} = \frac{1}{\alpha}S\left(\beta\right)\end{aligned} \tag{10}$$

from Equation 7. Thus, the maximum quasi-likelihood estimate will be the same as the maximum likelihood estimate from the Poisson model.
Having solved for $\hat{\beta}$, we have

$$\hat{\mu} = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right). \tag{11}$$

by Equation 2.31 of Wakefield's *Bayesian and Frequentist Regression Methods*, we can then compute

$$\hat{\alpha}_n = \frac{1}{n-2}\sum_{i=1}^{n}\frac{\left(y_i - \hat{\mu}_i\right)^2}{\hat{\mu}_i} \tag{12}$$

Then, the variance of our estimates is

$$\begin{aligned}\operatorname{Var}\left(\hat{\beta}\right) &\approx \hat{\alpha}_n\left(\hat{D}^{\intercal}\hat{V}^{-1}\hat{D}\right)^{-1} \\ &= \hat{\alpha}_n\left(\sum_{i=1}^{n}\begin{pmatrix}\hat{\mu}_i & x_i\hat{\mu}_i \\ x_i\hat{\mu}_i & x_i^2\hat{\mu}_i.\end{pmatrix}\right)^{-1} \\ &= \hat{\alpha}_n I_n\left(\hat{\beta}\right)^{-1} \end{aligned} \tag{13}$$

from Equation 8.

Sandwich Estimation

In sandwich estimation, we only need to specify an estimating function $G(\beta)$. Then, we can apply Equation 2.43 of Wakefield's *Bayesian and Frequentist Regression Methods* to compute the variance of our estimates:

$$\begin{aligned}\operatorname{Var}\left(\hat{\beta}\right) &= \frac{1}{n}\hat{A}^{-1}\hat{B}\left(\hat{A}^{-1}\right)^{\intercal} \\ \hat{A} &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}G\left(\hat{\beta}, X_i, Y_i\right) \\ \hat{B} &= \frac{1}{n}\sum_{i=1}^{n}G\left(\hat{\beta}, X_i, Y_i\right)G\left(\hat{\beta}, X_i, Y_i\right)^{\intercal}.\end{aligned}$$

We can reuse the score function from the quasi-likelihood estimate in Equation 10 without α, so

$$G\left(\hat{\beta}, X_i, Y_i\right) = \begin{pmatrix} Y_i - \hat{\mu}_i \\ X_i\left(Y_i - \hat{\mu}_i\right)\end{pmatrix} \tag{14}$$

Thus, our estimate for $\hat{\beta}$ will remain the same.
From Equations 8 and 13, we have that

$$\hat{A} = \frac{1}{n}\hat{D}\hat{V}^{-1}\hat{D} = \frac{1}{n}I_n\left(\hat{\beta}\right) \tag{15}$$

Length (mm)	0	1	2	3	4	5	6	7	8	9	10	11	12
1	2.247	2.640	2.842	2.908	3.099	3.126	3.245	3.328	3.355	3.383	3.572	3.581	3.681
10	1.901	2.132	2.203	2.228	2.257	2.350	2.361	2.396	2.397	2.445	2.454	2.454	2.474
20	1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958	1.966	1.997	2.006
50	1.339	1.434	1.549	1.574	1.589	1.613	1.746	1.753	1.764	1.807	1.812	1.840	1.852

Table 1: Failure stress data for four groups of fibers.

From Equation 10, we have that

$$\begin{aligned}
\hat{B} &= \frac{1}{n}\hat{D}^{\intercal}\hat{V}^{-1}\operatorname{diag}\left(RR^{\intercal}\right)\hat{V}^{-1}\hat{D} \\
&= \frac{1}{n}\hat{D}^{\intercal}\begin{pmatrix} \frac{(y_1-\hat{\mu}_1)^2}{\hat{\mu}_1^2} & & \\ & \ddots & \\ & & \frac{(y_n-\hat{\mu}_n)^2}{\hat{\mu}_n^2} \end{pmatrix}\hat{D} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^2}\begin{pmatrix} \hat{\mu}_i^2 & x_i\hat{\mu}_i^2 \\ x_i\hat{\mu}_i^2 & x_i^2\hat{\mu}_i^2 \end{pmatrix} \\
&= \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2\begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n}G\left(\hat{\beta},x_i,y_i\right)G\left(\hat{\beta},x_i,y_i\right)^{\intercal}. \qquad (16)
\end{aligned}$$

2. The data in Table 1 contain data on a typical reliability experiment and give the failure stresses (in GPa) of four samples of carbon fibers of lengths 1, 10, 20 and 50mm.

   (a) The exponential distribution $Y \mid \lambda \sim_{\text{iid}} \text{Exponential}(\lambda)$, is a simple model for reliability data:

$$p\left(y \mid \lambda\right) = \lambda\exp\left(-\lambda y\right), \qquad (17)$$

   with $\lambda, y > 0$. The hazard function is the probability of imminent failure and is given by

$$h\left(y \mid \lambda\right) = \frac{p\left(y \mid \lambda\right)}{S\left(y \mid \lambda\right)}, \qquad (18)$$

   where $S\left(y \mid \lambda\right) = \mathbb{P}\left(Y > y \mid \lambda\right)$ is the probability of failure beyond y. Derive the hazard function for the exponential distribution. Suppose we have a sample $y_1, \ldots, y_n$, of size n from an exponential distribution. Find the form of the MLE of λ and the asymptotic variance.

   Solution: The survival function can be derived with Equation 17 as

$$\begin{aligned}
S\left(y \mid \lambda\right) &= \mathbb{P}\left(Y > y \mid \lambda\right) \\
&= \int_y^{\infty}\lambda\exp\left(-\lambda t\right)\,\mathrm{d}t \\
&= -\exp\left(-\lambda t\right)\big|_y^{\infty} \\
&= \exp\left(-\lambda y\right). \qquad (19)
\end{aligned}$$

   With Equations 18 and 19, the hazard function is

$$h\left(y \mid \lambda\right) = \frac{p\left(y \mid \lambda\right)}{S\left(y \mid \lambda\right)} = \frac{\lambda\exp\left(-\lambda y\right)}{\exp\left(-\lambda y\right)} = \lambda. \qquad (20)$$

   Given $y_1, \ldots, y_n$, the log-likelihood function is

$$l\left(\lambda\right) = n\log\lambda - \lambda\sum_{i=1}^{n}y_i. \qquad (21)$$

Length (mm)	$\hat{\lambda}$	Standard error
1	0.317019	0.087925
10	0.432584	0.119977
20	0.571253	0.158437
50	0.599852	0.166369

Table 2: Results of fitting an exponential model for each length.

From Equation 21, the score function is

$$S(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} y_i. \tag{22}$$

Solving $S\left(\hat{\lambda}\right) = 0$, gives use the MLE, $\boxed{\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} y_i} = \frac{1}{\bar{y}}.}$

The asymptotic variance can be derived from the Fisher information

$$I_n\left(\hat{\lambda}\right) = \operatorname{Var}\left(S\left(\hat{\lambda}\right)\right) = \frac{n}{\hat{\lambda}^2}. \tag{23}$$

Thus, we have that $\boxed{\operatorname{Var}\left(\hat{\lambda}\right) = \frac{\hat{\lambda}^2}{n}.}$

(b) For each of the four groups in Table 1, estimate a separate λ, with an associated standard error. Examine the appropriateness of the exponential model via Q-Q plots.

Solution: The estimates and standard errors for λ are in Table 2.

From the Q-Q plots in Figure 1, we see that the exponential model is a very poor fit for the data. The sampled quantiles are very different than the theoretical quantiles. Thus, the model appears to be misspecified.

Code for calculations and plots can be found in `failure_stresses.ipynb`.

(c) Consider a quasi-likelihood approach to inference for λ under the model with

$$\begin{aligned} \mathbb{E}[Y \mid \lambda] &= \lambda^{-1} \\ \operatorname{Var}(Y \mid \lambda) &= \alpha\lambda^{-2} \end{aligned} \tag{24}$$

with $\alpha > 0$. Suggest an estimator for α. Estimate λ, α, and the standard errors, separately for each of the four groups in Table 1. What do the results suggest to you about the fit of the exponential model?

Solution: The quasi-likelihood score is

$$\begin{aligned} U(\lambda) &= D^{\intercal} V^{-1} \left(Y - \mathbb{E}[Y \mid \lambda]\right)/\alpha \\ &= \begin{pmatrix} -\lambda^{-2} & \cdots & -\lambda^{-2} \end{pmatrix} \begin{pmatrix} \lambda^2 & & \\ & \ddots & \\ & & \lambda^2 \end{pmatrix} \begin{pmatrix} \frac{Y_1 - \lambda^{-1}}{\alpha} \\ \vdots \\ \frac{Y_n - \lambda^{-1}}{\alpha} \end{pmatrix} \\ &= -\frac{1}{\alpha} \sum_{i=1}^{n} \left(Y_i - \lambda^{-1}\right) = -\frac{1}{\alpha} \left(\sum_{i=1}^{n} Y_i - n\lambda^{-1}\right). \end{aligned} \tag{25}$$

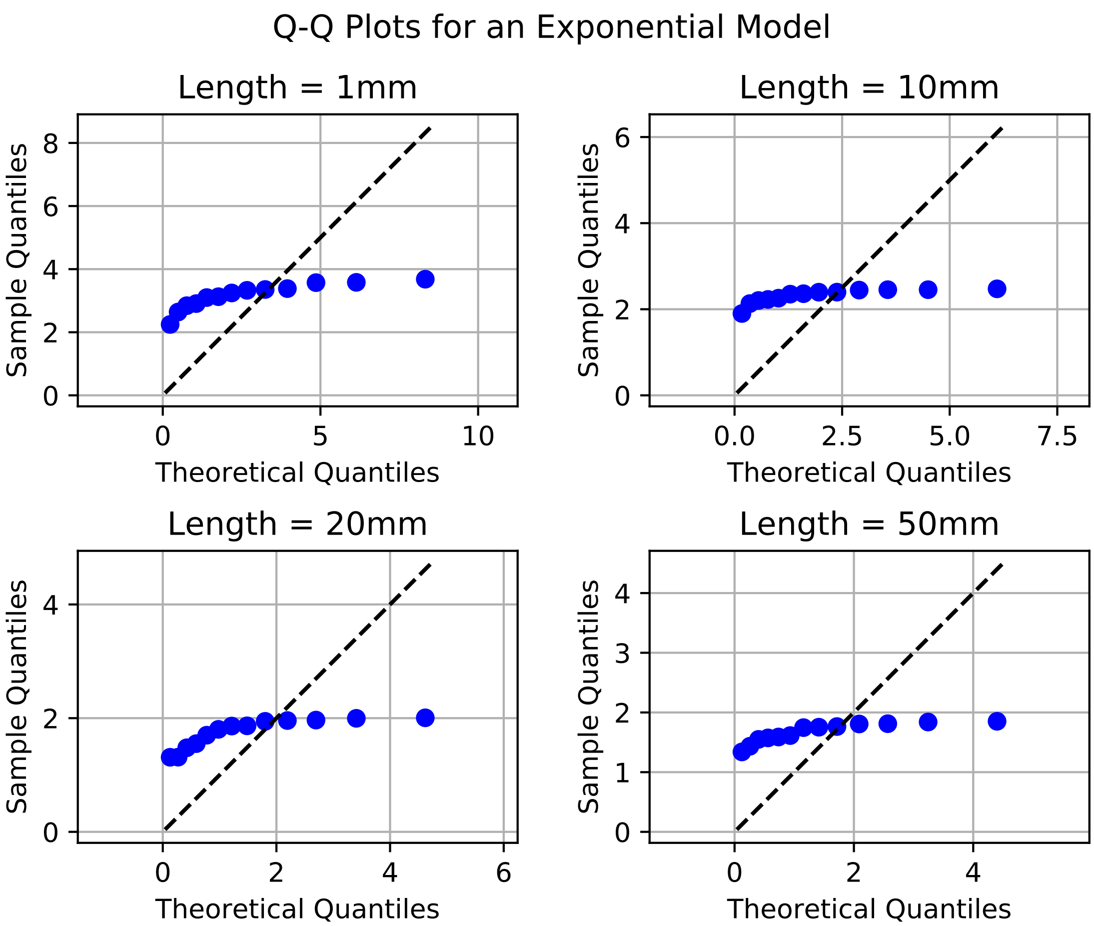

Figure 1: Q-Q plots at each length when fitting an exponential model using the MLE.

Length (mm)	$\hat{\lambda}$	$\hat{\alpha}$	Standard error
1	0.317019	0.016873	0.011421
10	0.432584	0.005078	0.008550
20	0.571253	0.021199	0.023068
50	0.599852	0.009633	0.016329

Table 3: Results of fitting a quasi-likelihood model for each length.

Solving Equation 25, $U\left(\hat{\lambda}\right) = 0$, we get $\hat{\lambda} = \bar{Y}^{-1}$, which is the same as the MLE estimate.
$\hat{\alpha}$ is given by Equation 2.31 of Wakefield's *Bayesian and Frequentist Regression Methods*:

$$\hat{\alpha} = \frac{1}{n-1}\sum_{i=1}^{n}\frac{\left(Y_i - \hat{\mu}\right)^2}{V\left(\hat{\mu}\right)} = \frac{\hat{\lambda}^2}{n-1}\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}^{-1}\right)^2. \tag{26}$$

We have that

$$\operatorname{Var}\left(U\left(\lambda\right)\right) = \mathbb{E}\left[-\frac{\partial U}{\partial \lambda}(\lambda)\right] = \frac{n\lambda^{-2}}{\alpha}, \tag{27}$$

we can estimate

$$\operatorname{Var}\left(\hat{\lambda}\right) = \operatorname{Var}\left(U\left(\hat{\lambda}\right)\right)^{-1} \approx \frac{\hat{\alpha}\hat{\lambda}^2}{n}, \tag{28}$$

which is the same as the variance for the MLE estimate multiplied by $\hat{\alpha}$. The results of fitting the quasi-likelihood model can be seen in Table 3.
From Equation 25, we see that a quadratic variance function leads to the same score function as Gamma distribution with fixed shape parameter α^{-1} and rate parameter $\lambda\alpha^{-1}$. $\alpha = 1$ would correspond to the exponential distribution, so it is unsurprising to see that our estimate for λ is the same as the MLE estimate.
Standard errors are much smaller than those estimated in Table 2. From Figure 1, we see that the residuals are underdispersed relative to an exponential model, so $\hat{\alpha} < 1$, which leads to the smaller standard error estimates.
Q-Q plots with the theoretical quantiles derived from Gamma $\left(\hat{\alpha}^{-1}, \lambda\hat{\alpha}^{-1}\right)$ in Figure 2. The points lie close the $y = x$ line. This suggests that the Gamma and quasi-likelihood model are more appropriate. They better capture the variance model compared to the exponential model.
Code for calculations and plots can be found in `failure_stresses.ipynb`.

(d) Obtain the form of the sandwich estimate for the variance of $\hat{\lambda}$. Numerically evaluate sandwich standard errors for the estimate of λ in each of the four groups.

Solution: We apply Equation 2.43 of Wakefield's *Bayesian and Frequentist*

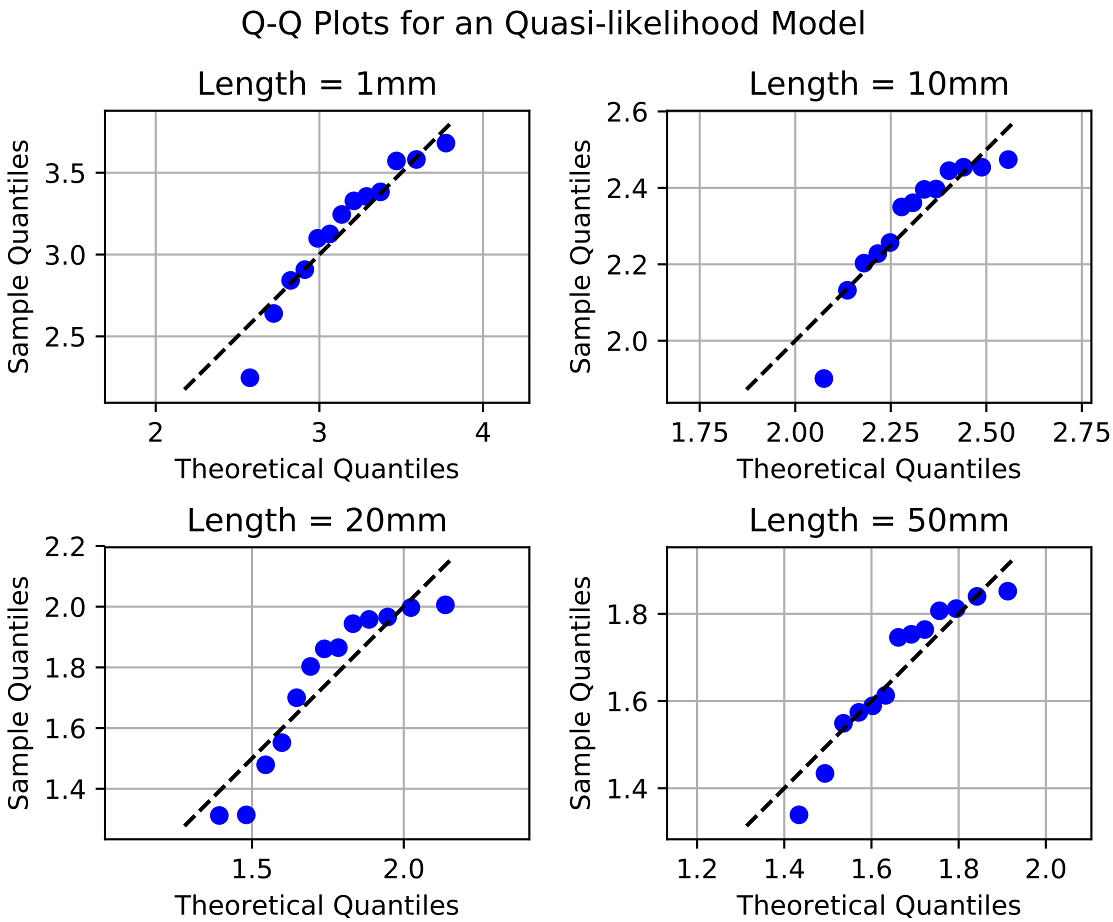

Figure 2: Q-Q plots at each length with a Gamma distribution. The shape parameter $\hat{\alpha}^{-1}$ and rate parameter $\hat{\lambda}\hat{\alpha}^{-1}$ were estimated using quasi-likelihood.

Length (mm)	$\hat{\lambda}$	Standard error
1	0.317019	0.010973
10	0.432584	0.008215
20	0.571253	0.022163
50	0.599852	0.015688

Table 4: Results of fitting a model with sandwich estimation for each length.

Regression Methods to compute the variance of our parameter estimate:

$$\begin{aligned}
\operatorname{Var}\left(\hat{\lambda}\right) &= \frac{1}{n}\hat{A}^{-1}\hat{B}\left(\hat{A}^{-1}\right)^{\intercal} \\
\hat{A} &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\lambda}G\left(\hat{\lambda}, Y_i\right) \\
\hat{B} &= \frac{1}{n}\sum_{i=1}^{n}G\left(\hat{\lambda}, Y_i\right)G\left(\hat{\lambda}, Y_i\right)^{\intercal},
\end{aligned}$$

where we resue the quasi-score from Equation 25 to specify

$$G\left(\lambda, Y_i\right) = \frac{1}{\lambda} - Y_i \tag{29}$$

as our estimating function.
Thus, we'll have that

$$\begin{aligned}
\hat{A} &= \frac{1}{n}I\left(\hat{\lambda}\right) = \frac{1}{\hat{\lambda}^2} \\
\hat{B} &= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \frac{1}{\hat{\lambda}}\right)^2.
\end{aligned} \tag{30}$$

Thus, our sandwich estimate will be

$$\boxed{\operatorname{Var}\left(\hat{\lambda}\right) = \frac{\hat{\lambda}^4}{n^2}\sum_{i=1}^{n}\left(Y_i - \frac{1}{\hat{\lambda}}\right)^2.} \tag{31}$$

The results of applying Equations 29 and 31 can be seen in Table 4.
The estimates for $\hat{\lambda}$ are of course the same as in Tables 2 and 3, since we reused the same score function. The standard errors are smaller than the exponential model since the data is underdispersed in that model. However, they are quite similar to those in the quasi-likelihood model despite not specifying a variance model. From Figure 2, we have evidence that the quasi-likelihood model fits the data well, so it is unsurprising that an empircal estimate would yield similar results.