HW 2 Solution

3.

For Gamma distribution, the variance is $Var(Y) = \mu^2/\nu$ and $V(\mu) = \mu^2$ (using the book notation). Then

$$A(y) = \int_{\infty}^{y} \frac{1}{V^{\frac{1}{3}}(\mu)} d\mu = \int_{\infty}^{y} \mu^{-\frac{2}{3}} d\mu = 3\mu^{\frac{1}{3}}$$

Hence,

$$Var(A(y)) = A'(\mu)\sqrt{Var(y)} = \mu^{\frac{1}{3}}/\nu$$

We obtain the Pearson residual for Gamma distribution is

$$r_p = \frac{y-\mu}{\sqrt{V(\mu)}} = \frac{y-\mu}{\mu/\sqrt{\nu}}$$

and the Anscombe residual is

$$r_a = \frac{3(y^{\frac{1}{3}} - \mu^{\frac{1}{3}})}{\sqrt{\mu^{\frac{1}{3}}/\nu}}$$

The R-code

```
in R, parameterized as
 The Gamma distribution with parameters 'shape' = a and 'scale' = s
     has density

                 f(x)= 1/(s^a Gamma(a)) x^(a-1) e^-(x/s)

     for x > 0, a > 0 and s > 0. The mean and variance are E(X) = a*s
     and Var(X) = a*s^2.
```

so $\mu = a * s$ and $\nu = a$

and $var(y) = a * s^2 = \mu^2/a = \mu^2/\nu$

fix scale=1, examine for different values of shape parameter

for small values, shape=1 or 2, Pearson residuals are considerably more skewed than Anscombe

for large values, shape=100 or 500, more similar

reason: as shape gets large, normal approx to gamma gets better

par(mfrow=c(1,2));

$y < -rgamma(500, shape =?, scale =?);$

#4.

$$f(z^*|y=1) = \frac{1}{2\pi}|\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(z^* - \mu_1)'\Sigma^{-1}(z^* - \mu_1)\}$$
$$f(z^*|y=2) = \frac{1}{2\pi}|\Sigma|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(z^* - \mu_2)'\Sigma^{-1}(z^* - \mu_2)\}$$

$$\Rightarrow f(z^*) = \sum_y f(z^*, y) = \sum_y f(z^*|y)f(y)$$
$$= f(z^*|y=1)f(y=1) + f(z^*|y=2)f(y=2)$$
$$= \frac{1}{2\pi}|\Sigma|^{-\frac{1}{2}}[\pi_1 \exp\{-\frac{1}{2}(z^* - \mu_1)'\Sigma^{-1}(z^* - \mu_1)\}$$
$$+ (1-\pi_1)\exp\{-\frac{1}{2}(z^* - \mu_2)'\Sigma^{-1}(z^* - \mu_2)\}]$$

$$\Rightarrow P(y=1|z^*) = \frac{f(z^*|y=1)f(y=1)}{f(z^*)}$$
$$= \frac{\pi_1 \exp\{-\frac{1}{2}(z^* - \mu_1)'\Sigma^{-1}(z^* - \mu_1)\}}{\pi_1 \exp\{-\frac{1}{2}(z^* - \mu_1)'\Sigma^{-1}(z^* - \mu_1)\} + (1-\pi_1)\exp\{-\frac{1}{2}(z^* - \mu_2)'\Sigma^{-1}(z^* - \mu_2)\}}$$

$$logit(P(y=1|z^*)) = log\frac{P(y=1|z^*)}{1-P(y=1|z^*)}$$
$$= log\frac{\pi_1 \exp\{-\frac{1}{2}(z^* - \mu_1)'\Sigma^{-1}(z^* - \mu_1)\}}{(1-\pi_1)\exp\{-\frac{1}{2}(z^* - \mu_2)'\Sigma^{-1}(z^* - \mu_2)\}}$$
$$= log\frac{\pi}{1-\pi} + \frac{1}{2}\mu_2'\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1'\Sigma^{-1}\mu_1 + (\mu_1 - \mu_2)'\Sigma^{-1}z^*$$
$$= \alpha + \beta z^*$$

5.

Let R_i be the unobserved true binary response for unit i with $\pi_i^* = P(R_i = 1)$ satisfying the logistic model

$$logit(\pi_i^*) = \beta^T x_i. \tag{3}$$

Suppose that the observed binary response, Y_i is subject to misclassification as follows,

$$P(Y_i = 1 \mid R_i = 0) = \delta_i$$
$$P(Y_i = 0 \mid R_i = 1) = \epsilon_i.$$

Suppose the misclassification errors satisfy the following condition

$$\frac{\delta_i}{\epsilon_i} = \frac{\pi_i^*}{1 - \pi_i^*}. \tag{4}$$

Under (4), can we consistently estimate β using the Y_i instead of R_i? [Hint: Just determine the relationship between $logit(\pi_i)$ and $\beta^T x_i$, where $\pi_i = P(Y_i = 1)$]. Does (4) seem like a plausible condition? Comment. Finally, would you expect the variance of $\hat{\beta}$ to compare if estimating it using Y_i versus R_i? Explain.

$$\pi_i = P(Y_i = 1) = \sum_{r=0}^{1} P(Y_i = 1 \mid R_i = r)\, P(R_i = r)$$

$$= \delta_i (1 - \pi_i^*) + (1 - \epsilon_i)\pi_i^*$$

using (4)
$$= \epsilon_i \frac{\pi_i^*}{1-\pi_i} (1 - \pi_i^*) + (1 - \epsilon_i)\pi_i^*$$

$$= \pi_i^*$$

→ so can consistently estimate β using Y_i instead of R_i.

- Is (4) plausible? may not be. says misclassification probability is different for every i and have same relationship (4) for every i.
- it is the same. First, can see from $Y_i \sim Be(\pi_i^*)$, same as R_i. Also if derive it, after algebra

$$Var(Y_i) = E[Var(Y_i \mid R_i)] + Var[E(Y_i \mid R_i)] = \pi_i^*(1 - \pi_i^*$$