BIOS 7345: Advanced Regression Analysis I

Andrew J. Spieker, Ph.D.

Assistant Professor of Biostatistics
Vanderbilt University Medical Center

Set 15: Overdispersion and quasi-likelihood

Version: 09/17/2023

Table of Contents

Variance estimators (so far)

Recall: Likelihood-based variance

- If there is no nuisance parameter, we estimate the variance as:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}.$$

- If there is a nuisance parameter, we estimate the variance as:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = \widehat{\phi}(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1} = \left(\frac{1}{N-K} \sum_{i=1}^{N} \frac{(y_i - \widehat{\mu}_i)^2}{V(\widehat{\mu}_i)} \right) (\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}.$$

- $\mathbb{A}_n(\widehat{\boldsymbol{\beta}})$ is proportional to the (expected) Fisher information (for $\boldsymbol{\beta}$).
- These estimators are valid if the likelihood is completely correct.
- We argued previously that these estimators are also valid if both the mean model and the mean-variance relationship implied by the GLM are correctly specified.

Variance estimators (so far)

Recall: Sandwich-based variance (version 1)

- The following estimator protects against both mean model and mean-variance misspecification:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}))^{-1}(\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}))(\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}))^{-1}.$$

- If the mean model is incorrect, this estimator will not be valid if **X** is fixed (i.e., does not vary by study replicate).

Variance estimators (so far)

Recall: Sandwich-based variance (version 2)

- The following estimator protects against misspecification of the mean-variance relationship:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] = (\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}(\mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}))(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}.$$

- If the mean model is incorrect, this estimator will not be valid if $\mathbf{X}$ is fixed (i.e., does not vary by study replicate).
- When the canonical link is used, $\mathbb{A}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}})$ and $\mathbb{A}_N(\widehat{\boldsymbol{\beta}})$ are the same; therefore, this estimator would be valid under mean model misspecification under the canonical link.

Variance estimators (so far)

Recall: Bootstrap variance (version 1)

- We discussed a flavor of the bootstrap procedure that re-samples "independent rows" from the original data (with replacement).
- This is consistent with a sampling mechanism in which the design matrix changes from replicate to replicate.
- If the mean model is incorrect and $\mathbf{X}$ is really fixed by design, this will tend to overstate the variance.

VARIANCE ESTIMATORS (SO FAR)

Recall: Bootstrap variance (version 2)

- We discussed a flavor of the bootstrap procedure that holds the design matrix fixed and re-samples (or re-generates) outcomes only.
- This is consistent with a sampling mechanism in which the design matrix is fixed by design.
- If the mean model is incorrect and $\mathbf{X}$ is really random by design, this will tend to understate the variance.

Variance estimators (so far)

Next step: Quasi-likelihood

- These notes introduce quasi-likelihood; in my mind, it is easier to first discuss a specific application of quasi-likelihood (overdispersion, in particular) and then go through the framework a little more thoroughly.

Table of Contents

Overdispersion

Example: Poisson regression

- Imagine we are modeling a count outcome, Y_i, as a function of $\mathbf{X}_i$.
- A natural GLM: $Y_i \sim \text{Poisson}(\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}))$.
- Here is the salient information implied by the GLM:
  - Mean model: $\log(\text{E}[Y_i|\mathbf{x}]) = \mathbf{x}^T \boldsymbol{\beta}$.
  - Mean-variance relationship: $\text{Var}[\mathbf{y}] = \boldsymbol{\mu}$.
- The Poisson family is a single-parameter family; it does not have a nuisance parameter.
- What if I believe the mean model to be correct, but *not* the mean-variance relationship?
  - Any of the many variance estimators we have discussed so far should be valid *except* the likelihood-based one. Why?

Example: Overdispersed/underdispersed Poisson

- What if I believe the mean model, and that $\text{Var}[Y_i|\mathbf{x}_i] = \varphi\mu_i$?
- We refer to φ as the *dispersion* parameter:
  - If $0 < \varphi < 1$, we say there is underdispersion.
  - If $0 < \varphi < \infty$, we say there is overdispersion.
- Remember: no nuisance parameter in the Poisson family. Believing the mean-variance relationship implied by the Poisson family means you believe that $\text{Var}[Y_i|\mathbf{x}_i] = \mu_i$ exactly.
- Believing the mean-variance relationship "up to a fixed, unknown proportionality constant" effectively introduces (enforces) a nuisance where it doesn't exist in the single-parameter likelihood.
  - This is, as we will see, an application of a broader framework referred to as quasi-likelihood.
- Appropriate variance estimator in this case:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] \quad = \quad \widehat{\varphi}(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1} = \left(\frac{1}{N-K}\sum_{i=1}^{N}\frac{(y_i-\widehat{\mu}_i)^2}{V(\widehat{\mu}_i)}\right)(\mathbb{A}_N(\widehat{\boldsymbol{\beta}}))^{-1}.$$

Example: Overdispersed/underdispersed Poisson

- I leave it to you to verify the following estimating equations for Poisson regression (log-link):

$$\mathbf{X}^T(\mathbf{y} - \exp(\mathbf{x}_i^T \boldsymbol{\beta})) = \mathbf{0}$$

- I leave it to you to verify the following components of the sandwich:

$$\begin{aligned} \mathbb{A}_N(\widehat{\boldsymbol{\beta}}) &= \mathbf{X}^T \text{diag}(\exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}}))\mathbf{X}. \\ \mathbb{B}_N^{\text{obs}}(\widehat{\boldsymbol{\beta}}) &= \mathbf{X}^T \text{diag}(y_i - \exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}}))^2 \mathbf{X}. \end{aligned}$$

- I leave it to you to verify the estimator of the dispersion parameter:

$$\widehat{\varphi} = \frac{1}{N-K} \sum_{i=1}^{N} \frac{(y_i - \exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}}))^2}{\exp(\mathbf{x}_i^T \widehat{\boldsymbol{\beta}})}.$$

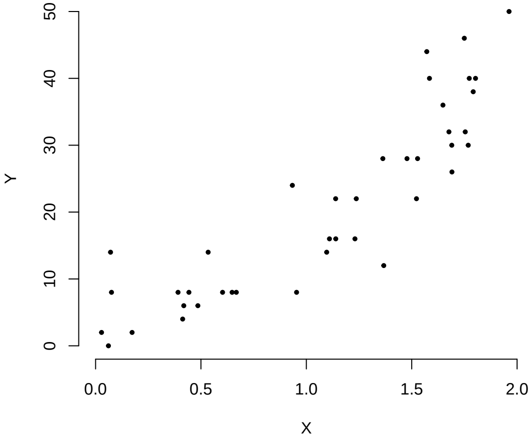
Example: Overdispersed Poisson

- Consider generating $Y = 2Y^*$, where $Y^* \sim \text{Poisson}(\exp(\mathbf{x}^T \boldsymbol{\beta}^*))$.
- Then, $\text{E}[Y|\mathbf{x}] = \exp(\mathbf{x}^T \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_0^* + \log(2), \beta_1^*, \ldots, \beta_K^*)$. The Poisson mean model is correctly specified.
- Further, $\text{Var}[Y|\mathbf{x}] = 4\text{Var}[Y^*|\mathbf{x}] = 2\text{E}[Y|\mathbf{x}] \Rightarrow \varphi = 2$. The Poisson mean-variance relationship is not correctly specified, but it is correct up to a proportionality constant that can be estimated and taken into account.

Example: Overdispersed Poisson

```
## Set seed for reproducibility
set.seed(7345)

## Set "true" value of beta
beta <- matrix(c(1, 1), nrow = 2)

## Generate data
n <- 40
X <- matrix(cbind(1, runif(n, 0, 2)), ncol = 2)
y <- 2*rpois(n, exp(X %*% beta))

## The true value of beta is really (1 + log(2), 1).
```

Example: Overdispersed Poisson



Overdispersion

Example: Overdispersed Poisson

- I leave it to you to verify that $\widehat{\beta}_1 = 1.22$.
- I leave it to you to verify that $\widehat{\varphi} = 1.94$.
- I leave it to you to verify the following standard errors:
  1. Likelihood-based (invalid): $\widehat{\text{SE}}[\widehat{\beta}_1] = 0.0764$.
  2. Sandwich-based (valid): $\widehat{\text{SE}}[\widehat{\beta}_1] = 0.112$.
  3. Non-parametric bootstrap (valid): $\widehat{\text{SE}}[\widehat{\beta}_1] \approx 0.115$.
  4. Quasi-Poisson (valid): $\widehat{\text{SE}}[\widehat{\beta}_1] = 0.106$.
- Once you know how to program a GLM in R, going the extra step to estimate the dispersion parameter is straightforward.
- In R, you can use the following family specification to account for overdispersion (or underdispersion):

```
glm(..., family = quasipoisson(link = "log"))
```

Example: Overdispersed Poisson

- With the previous example, we just motivated (and provided a *very* common application of) quasi-likelihood theory.
- We previously made the argument that the score equations for GLMs are uniquely determined by the first and second moments of the corresponding likelihood; even then, there is typically an asymptotically valid estimation procedure for the variance even if neither moment is correct (there are many considerations that go into selecting the right variance estimator).
- We'll make this argument again within the quasi-likelihood framework, and discuss the implications in a little more detail.

Table of Contents

Quasi score:

- Consider that the quantity

$$U(\mu; Y) = \frac{Y - \mu}{\varphi V(\mu)}$$

  behaves in many ways like a score function.
- The function $V(\cdot)$ specifies the mean-variance relationship.
- I leave it to you to verify that:
  - $\mathrm{E}[U] = 0$.
  - $\mathrm{Var}[U] = (\varphi V(\mu))^{-1}$.
  - $-\mathrm{E}[\partial U / \partial \mu] = (\varphi V(\mu))^{-1}$.

Quasi-(log)likelihood:

- Therefore, it stands to reason that the quantity

$$\sum_{i=1}^{N} Q(\mu_i, y_i) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\varphi V_i(t)} dt,$$

  when it exists, behaves much like a log-likelihood function.
- If it does exist, its derivative (with respect to $\boldsymbol{\beta}$) takes the form of the estimating equations we have been considering:

$$\mathbf{D}^T \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})/\varphi = \mathbf{0}.$$

Quasi-(log)likelihood:

- The quasi-(log)likelihood looks something like the log-likelihood of the exponential families we've been discussing:

$$\sum_{i=1}^{N} Q(\mu_i, y_i) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\varphi V_i(t)} dt.$$

- The specific form of the quasi-(log)likelihood (as a function of $\boldsymbol{\mu}$) is determined by specifying a function, $V(\cdot)$, that corresponds to a mean-variance relationship.
- There are certain things we might expect if we were to choose specific forms of $V(\cdot)$. Let's try a couple!

Quasi-(log)likelihood: Normal

- If we choose $V(\mu) = 1$, we expect the quasi-(log)likelihood to resemble the log-likelihood of a normal distribution.

$$\begin{aligned} Q(\mu, y) &= \int_y^\mu \frac{y - t}{\varphi} dt \\ &= \frac{1}{\varphi} \int_y^\mu (y - t)\, dt \\ &= -\frac{1}{2\varphi}(y - \mu)^2. \end{aligned}$$

- Important notes:
  - This contains only the salient information from the normal likelihood.
  - The dispersion parameter, φ, serves the role of σ^2.

Quasi-likelihood

Quasi-(log)likelihood: Poisson

- If we choose $V(\mu) = \mu$, we expect the quasi-(log)likelihood to resemble the log-likelihood of a Poisson distribution.

$$
\begin{aligned}
Q(\mu, y) &= \int_y^\mu \frac{y - t}{\varphi t} dt \\
&= \frac{1}{\varphi} \int_y^\mu \left(y t^{-1} - 1 \right) dt \\
&= \frac{1}{\varphi} \left(y \log(\mu) - \mu - y(1 + \log(y)) \right).
\end{aligned}
$$

- Important notes:
  - Restrictions: $\mu > 0$; $y \geq 0$.
  - $c(y) = -y(1 + \log(y))$ is constant with respect to μ and is irrelevant in deriving the quasi-score for $\boldsymbol{\beta}$.
  - The Poisson distribution does not have a nuisance parameter; this framework introduces overdispersion and underdispersion to a model in which we believe the mean and variance are directly proportional.

Quasi-(log)likelihood: Bernoulli

- If we choose $V(\mu) = \mu(1-\mu)$, we expect the quasi-(log)likelihood to resemble the log-likelihood of a Bernoulli distribution.

$$
\begin{aligned}
Q(\mu, y) &= \int_y^\mu \frac{y-t}{\varphi t(1-t)} dt \\
&= (\ldots \text{Informal homework!})
\end{aligned}
$$

- Important notes:
  - Be mindful of restrictions on μ and y.
  - If Bernoulli observations are independent, it is theoretically impossible to have underdispersion or overdispersion under a correct mean model.
  - If there *is* underdispersion or overdispersion in this case, it must originate from correlation between the observations (sneak preview of BIOS 7346 material).

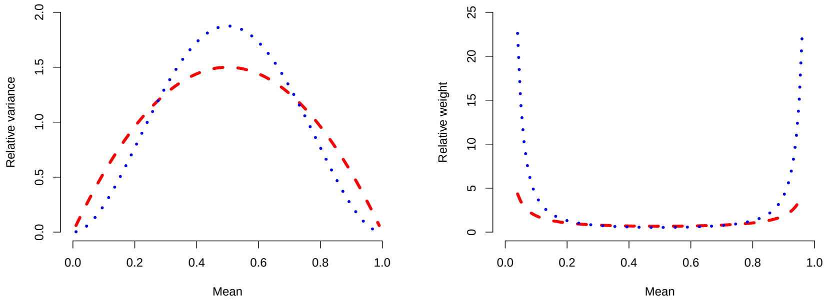
Quasi-(log)likelihood: ???

- If we choose $V(\mu) = \mu^2(1-\mu)^2$, we expect the quasi-(log)likelihood to resemble the log-likelihood of a. . . ???

$$Q(\mu, y) = \int_y^\mu \frac{y-t}{\varphi t^2(1-t)^2} dt.$$

- Important notes:
  - This choice of $V(\mu)$ does not originate from a probability function.
  - "Quasi-likelihood" is so termed because the resulting estimating equations need not originate from an actual likelihood.
- What might be a motivation to use this choice of a variance function for an outcome bounded between zero and one?

Implied weights: $V_1(\mu) \propto \mu(1-\mu)$ vs. $V_2(\mu) \propto \mu^2(1-\mu)^2$



GLMs with a nuisance parameter: Really just quasi-likelihood

- Quasi-likelihood is, in my opinion, the best way to think about a GLM. Thinking about GLMs in this framework involves the following steps:
  1. Specify a mean model, $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$.
  2. Specify a mean-variance relationship, $\mathbf{V}(\boldsymbol{\mu})$.
  3. Select an appropriate variance estimation procedure based on the following considerations (some of which interact with one another).
     - ⋆ Do you believe the mean model and/or mean-variance relationship?
     - ⋆ Might there by overdispersion or underdispersion?
     - ⋆ Is the link function "canonical" given the mean-variance relationship?
     - ⋆ Is $\mathbf{X}$ fixed by design or random by design?
     - ⋆ What is your sample size? What is your computing power?
- This gives you some flexibility. For instance, when thinking within this framework, you don't need to have count data to justify fitting "Poisson regression," if you believe the mean model and mean-variance relationship *implied* by a Poisson regression model.

Optimality:

- Consider the mean model $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$ (assume correctly specified).
- Consider the following linear unbiased estimating equations to estimate $\boldsymbol{\beta}$:

$$\mathbf{D}^T \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0},$$

  where $\mathbf{W}$ is a user-specified weight matrix that may depend upon $\boldsymbol{\beta}$.
- If $\mathbf{W} \propto \mathbf{V}^{-1}$, the resulting estimator $\widehat{\boldsymbol{\beta}}$ will have the smallest variance of all estimators arising from linear unbiased estimating equations.
- This is a statement (provided without proof) of the Gauss-Markov theorem as it applies to quasi-likelihood (and by extension, GLMs). You've already seen several variants of this theorem (four, I think?) for linear regression.

Another application: Robust estimation

- Quasi-likelihood has other applications that are outside the scope of this course.
- One other popular application is to robust estimation, in which you select the weights of a model in a way that *bounds* the influence of an observation (the influence of an observation grows without bound in a regression model when they are equally weighted).

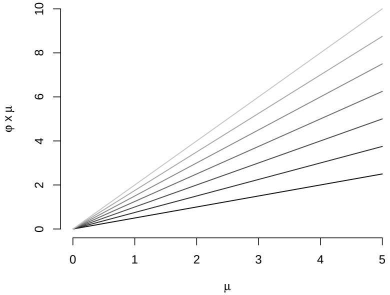
Table of Contents

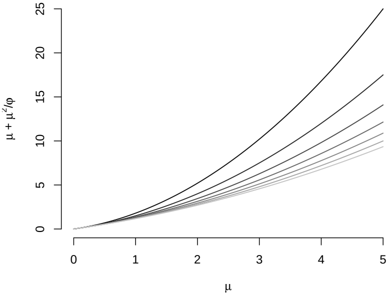
Negative binomial regression

Poisson regression: Variance choices

- To finish off this set of slides, let's talk about *one* more way to deal with overdispersion (overdispersed count data, in this case).
- As we know, count data are often analyzed using Poisson regression.
  - If we believe the mean model and that $\text{Var}[Y_i|\mathbf{x}_i] = \text{E}[Y_i|\mathbf{x}_i]$, we should be justified in the likelihood-based approach.
  - If we believe the mean model and that $\text{Var}[Y_i|\mathbf{x}_i] = \varphi\text{E}[Y_i|\mathbf{x}_i]$ for some unknown φ, we can use quasi-Poisson regression.
  - If we believe the mean model but not that $\text{Var}[Y_i|\mathbf{x}_i] = \varphi\text{E}[Y_i|\mathbf{x}_i]$, we can employ bootstrap or sandwich-based techniques.
- Even the condition $\text{Var}[Y_i|\mathbf{x}_i] = \varphi\text{E}[Y_i|\mathbf{x}_i]$ is restrictive. One alternative is to allow $\text{Var}[Y_i|\mathbf{x}_i] = \text{E}[Y_i|\mathbf{x}_i] + \text{E}[Y_i|\mathbf{x}_i]^2/\varphi$ for an unknown φ.
- This doesn't quite fit into our quasi-likelihood framework. Why?

Modeling overdispersion: Different choices of φ





Poisson regression: Variance choices

- Both the parametric and the quasi-likelihood formulation of a GLM result in estimating equations that take the following form:

$$\begin{aligned} \mathbf{D}^T(\boldsymbol{\beta})\mathbf{V}^{-1}(\boldsymbol{\beta})(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}))/\phi &= \mathbf{0} \quad \text{(likelihood)} \\ \mathbf{D}^T(\boldsymbol{\beta})\mathbf{V}^{-1}(\boldsymbol{\beta})(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}))/\varphi &= \mathbf{0} \quad \text{(quasi-likelihood)} \end{aligned}$$

- The value of the nuisance/dispersion parameter does not have any influence on the solution to the estimating equations $\widehat{\boldsymbol{\beta}}$. Why?
  - For the special exponential family we've been considering for GLMs, recall that $\text{Var}[Y_i|\mathbf{x}_i] = \phi b''(\theta_i)$ is linear in the nuisance parameter, ϕ.
  - For quasi-likelihood, we assume that $\text{Var}[Y_i|bx_i] = \varphi V(\mu_i)$ is linear in the dispersion parameter, φ.
- If we assume $\text{Var}[Y_i|\mathbf{x}_i] = \mu_i + \mu_i^2/\varphi$, then the estimating equations take the form: $\mathbf{D}^T(\boldsymbol{\beta})\mathbf{V}^{-1}(\boldsymbol{\beta}, \varphi)(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta})) = \mathbf{0}$, whereby the solution to the estimating equations, $\widehat{\boldsymbol{\beta}}$, *does* depend upon φ.

Negative binomial regression

Variance function: $\text{Var}[Y_i|\mathbf{x}_i] = \mu_i + \mu_i^2/\varphi$

- Natural question: what are the origins of this variance function?
- I confess: I answer this begrudgingly because the answer is a parametric model, in part undermining the point I'm trying to make in this set of notes, which is that we don't need a likelihood to motivate a suitable estimating equation.
- Nevertheless, context can be helpful, so here's the scoop.
- Suppose $Z \sim \text{Gamma}(\varphi, \varphi)$ is a latent variable (so that $\text{E}[Z] = 1$ and $\text{Var}[Z] = 1/\varphi$), and let $Y|Z \sim \text{Poisson}(\mu Z)$. It is straightforward to show that the *marginal* probability function for Y is given by:

$$p_Y(y; \mu, \varphi) = \frac{\Gamma(\varphi + y)}{\Gamma(\varphi) y!} \frac{\varphi^{\varphi} \mu^y}{(\mu + \varphi)^{\varphi + y}},$$

  such that $Y \sim \text{NegativeBinomial}(k = \varphi, p = \mu/(\mu + \varphi))$.

Negative binomial distribution:

- Let's investigate $Y \sim \text{NegativeBinomial}(k = \varphi, p = \mu/(\mu + \varphi))$, which originates from counting the number of failures in a sequence of i.i.d. Bernoulli trials until the first φ successes are achieved.
    - We're not using this parametric family because of its interpretation in this way, but because of the implied mean-variance relationship that we seek to derive.
- If φ is known, this takes the exponential form appropriate for a GLM.

$$p_Y(y; \mu, \varphi) = \exp\left(y \log\left(\frac{\mu}{\mu + \varphi}\right) - \varphi \log(\mu + \varphi) + c(y, \varphi) \right).$$

- If φ is unknown, this is not suitable for the GLM framework as we have discussed it. Let's start by treating φ as known.

Negative binomial regression

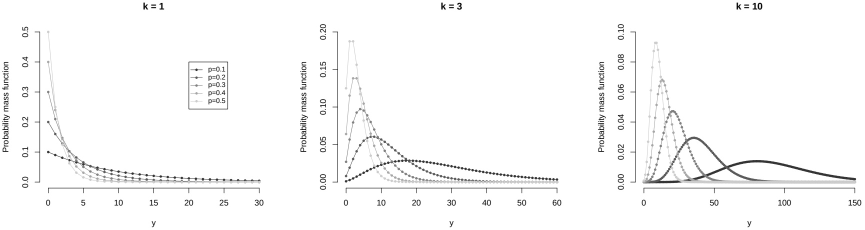
Negative binomial distribution:

- Mass function:

$$p_Y(y;\mu,\varphi) \;=\; \exp\left(y\log\left(\frac{\mu}{\mu+\varphi}\right)-\varphi\log(\mu+\varphi)+c(y,\varphi)\right).$$

- I leave it to you to show the following:
  - $\theta = \log(\mu/(\mu+\varphi))$ (canonical parameter).
  - $b(\theta) = \varphi\log(\varphi/(1-\exp(\theta)))$.
  - $b'(\theta) = \varphi\exp(\theta)/(1-\exp(\theta)) = \mu$.
  - $V(\mu) = \mu + \mu^2/\varphi$ (this is the punchline).
- We rarely use the canonical link for negative binomial regression.
- Rather, this model is a convenient way of modeling count data in which there is overdispersion with respect to the Poisson model.
- Note that the negative binomial distribution resembles a Poisson distribution as φ grows (reflected by the mean-variance relationship).

Negative binomial distribution: Mind the different axis scales



Negative binomial regression: Unknown φ

- If φ is known, estimation of $\boldsymbol{\beta}$ is straightforward under the likelihood framework or the framework of estimating equations.
- The estimating equations for $\boldsymbol{\beta}$ unavoidably depend upon φ, but we can estimate it iteratively with $\boldsymbol{\beta}$.
- I leave it to you to verify the following estimating equations for negative binomial regression under the log link:

$$\mathbf{X}^T \text{diag}\left(\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{\exp(\mathbf{x}_i^T\boldsymbol{\beta}) + \exp(\mathbf{x}_i^T\boldsymbol{\beta})^2/\varphi}\right)(\mathbf{y} - \exp(\mathbf{X}\boldsymbol{\beta})) = \mathbf{0}.$$

- I also leave it to you to verify the following method-of-moments estimator for φ given $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\varphi} = \left[\frac{1}{N-K}\sum_{i=1}^{N}\frac{(y_i - \exp(\mathbf{x}_i^T\widehat{\boldsymbol{\beta}}))^2 - \exp(\mathbf{x}_i^T\widehat{\boldsymbol{\beta}}))}{\exp(\mathbf{x}_i^T\widehat{\boldsymbol{\beta}})^2}\right]^{-1}$$

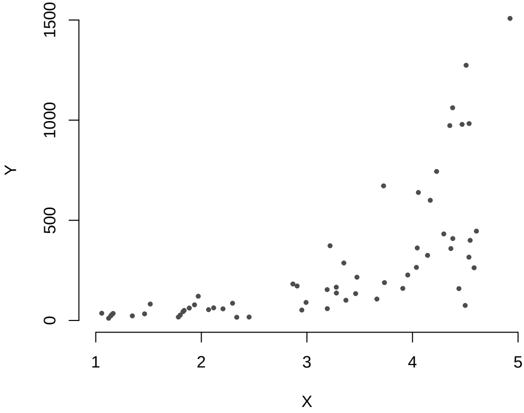
Negative binomial regression: Unknown φ

- This suggests iterating estimation of $\boldsymbol{\beta}$ and φ using, e.g., the Gauss-Newton algorithm. There are technical/theoretical details surrounding why you can do this; I am putting those aside for the sake of time.
- The function `glm.nb()` in R will not iterate estimation; it will treat φ as fixed after estimating it in a single step (asymptotically valid).

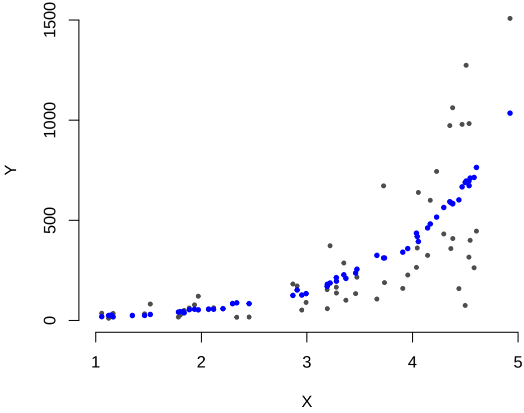
Negative binomial regression: Example

```
## Set seed for reproducibility
set.seed(7345)

## Set sample size
n <- 60

## Set true parameters
beta <- c(2, 1)
phi <- 3

## Generate data
X <- cbind(1, runif(n, 1, 5))
prob <- phi/(phi + exp(X %*% beta))
y <- rnbinom(n, size = phi, prob = prob)

## Initialize beta
betaj <- c(1, 1)
phij <- 1

## Set tolerance and count iterations
tol <- 1
iter <- 1
```

Negative binomial regression: Example



Negative binomial regression: Comparison to `rpois()`



Negative binomial regression: Example

```
while(tol > 1e-20 & iter < 50)
{
  ## Store prior iteration
  betaj.prior <- betaj

  ## Linear predictor
  muj <- c(exp(X %*% betaj))

  V <- diag(muj + muj^2/phij)

  ## Estimating function
  Gn <- t(X) %*% diag(muj) %*% solve(V) %*% (y - muj)

  ## An
  An <- t(X) %*% diag(muj^2) %*% solve(V) %*% X

  ## Update to next iteration
  betaj <- betaj + solve(An) %*% Gn

  ## Nuisance
  phij <- ((1/(n - 2))*sum(((y - muj)^2 - muj)/(muj^2)))^(-1)

  ## Determine if conditions are met
  tol <- sum((betaj - betaj.prior)^2)
  iter <- iter + 1
}
```

Negative binomial regression: Results

```
> iter
[1] 12

> tol
[1] 2.42323e-23

## Estimate
> betaj
          [,1]
[1,] 2.1642150
[2,] 0.9429575

## Standard errors
> sqrt(diag(solve(An)))
[1] 0.21660667 0.06387519
```

Negative binomial regression: Comparison to Poisson regression

```
zz2 <- glm(y ~ X[,2], family = poisson(link = "log"))

## Estimate (consistent)
> summary(zz2)$coef[,1]
(Intercept)      X[, 2]
   1.892569    1.015723

## Model-based standard errors (invalid!)
> sqrt(diag(vcov(zz2)))
(Intercept)      X[, 2]
 0.04549637  0.01097969

## Sandwich standard errors (valid!)
> sqrt(diag(sandwich(zz2)))
(Intercept)      X[, 2]
  0.3621749   0.1013718
```

Negative binomial regression: Comparison to quasi-Poisson

```
zz3 <- glm(y ~ X[,2], family = quasipoisson(link = "log"))

## Estimate (consistent; same as Poisson GLM)
 > summary(zz3)$coef[,1]
(Intercept)      X[, 2]
   1.892569    1.015723

## Dispersion parameter (not consistent for phi; different model!)
> summary(zz3)$dispersion
[1] 98.56069

## Model-based standard errors (invalid, but "better than model-based.")
> sqrt(diag(vcov(zz3)))
(Intercept)      X[, 2]
  0.4516777   0.1090039
```

Summary

So far:

- Overdispersion and quasi-likelihood.

Summary

Up next:

- Hypothesis testing for GLMs.