

Generalized Linear Models

Anton Korobeynikov

March 15, 2016

Motivation for nonlinear models

The key properties of a linear model:

$$E(Y|X) = X\beta \quad \text{and} \quad \text{var}(Y|X) \propto I.$$

What to do when these conditions are not met?

Motivation for nonlinear models

The key properties of a linear model:

$$E(Y|X) = X\beta \quad \text{and} \quad \text{var}(Y|X) \propto I.$$

What to do when these conditions are not met?

- ▶ Use robust methods.
- ▶ We can transform Y so that the linear model assumptions are approximately satisfied.

However it is often difficult to find a transformation that simultaneously linearizes the mean and gives constant variance.

Motivation for nonlinear models

More problems:

- ▶ If Y lies in a restricted domain (e.g. $Y = 0, 1$ or $Y = 0, 1, \dots$), parameterizing $E(Y|X)$ as a linear function of X violates the domain restriction.

Motivation for nonlinear models

More problems:

- ▶ If Y lies in a restricted domain (e.g. $Y = 0, 1$ or $Y = 0, 1, \dots$), parameterizing $E(Y|X)$ as a linear function of X violates the domain restriction.

Generalized linear models (GLM's) are a class of nonlinear regression models that can be used in certain cases where linear models are not appropriate.

Motivating Example: Logistic regression

Logistic regression is used for binary outcome data, where $Y = 0$ or $Y = 1$. It is defined by the probability mass function

$$P(Y = 1|X = x) = \frac{\exp(\beta^T x)}{1 + \exp(\beta^T x)} = \frac{1}{1 + \exp(-\beta^T x)},$$

which implies that

$$P(Y = 0|X = x) = 1 - P(Y = 1|X = x) = \frac{1}{1 + \exp(\beta^T x)},$$

where $x_0 = 1$ so β_0 is the intercept.

Logistic regression

The **logit function**

$$\text{logit}(x) = \log(x/(1 - x))$$

maps the unit interval onto the real line. The **inverse logit function**, or **expit function**

$$\text{expit}(x) = \text{logit}^{-1}(x) = \frac{\exp(x)}{1 + \exp(x)}$$

maps the real line onto the unit interval.

In logistic regression, the logit function is used to map the **linear predictor** $\beta^T X$ to a probability.

Logistic regression

The linear predictor in logistic regression is the **conditional log odds**:

$$\log \left[\frac{P(Y = 1|X)}{P(Y = 0|X)} \right] = \beta^T X.$$

Logistic regression

The linear predictor in logistic regression is the **conditional log odds**:

$$\log \left[\frac{P(Y = 1|X)}{P(Y = 0|X)} \right] = \beta^\top X.$$

Thus one way to interpret a logistic regression model is that a one unit increase in X_j results in a change of β_j in the conditional log odds.

Or, a one unit increase in X_j results in a multiplicative change of $\exp(\beta_j)$ in the conditional odds.

Latent variable model for logistic regression

It may make sense to view the binary outcome Y as being a dichotomization of a latent continuous outcome Y_c ,

$$Y = \mathbb{I}(Y_c \geq 0).$$

Suppose $Y_c|X$ follows a logistic distribution, with CDF

$$F(Y_c|X) = \frac{\exp(Y_c - \beta^\top X)}{1 + \exp(Y_c - \beta^\top X)}.$$

Latent variable model for logistic regression

It may make sense to view the binary outcome Y as being a dichotomization of a latent continuous outcome Y_c ,

$$Y = \mathbb{I}(Y_c \geq 0).$$

Suppose $Y_c|X$ follows a logistic distribution, with CDF

$$F(Y_c|X) = \frac{\exp(Y_c - \beta^\top X)}{1 + \exp(Y_c - \beta^\top X)}.$$

In this case, $Y|X$ follows the logistic regression model:

$$P(Y = 1|X) = P(Y_c \geq 0|X) = 1 - \frac{\exp(0 - \beta^\top X)}{1 + \exp(0 - \beta^\top X)} = \frac{\exp(\beta^\top X)}{1 + \exp(\beta^\top X)}.$$

Mean/variance relationship for logistic regression

Since the mean and variance of a Bernoulli trial are linked, the mean structure

$$E(Y|X) = P(Y = 1|X) = \frac{\exp(\beta^T X)}{1 + \exp(\beta^T X)}$$

also determines the variances

$$\text{var}(Y|X) = P(Y = 1|X) \cdot P(Y = 0|X) = \frac{1}{2 + \exp(\beta^T X) + \exp(-\beta^T X)}.$$

Since the variance depends on X , logistic regression models are always heteroscedastic.

Logistic regression and case-control studies

Suppose we sample people based on their disease status D ($D = 1$ is a **case**, $D = 0$ is a **control**).

We are interested in a binary marker $M \in \{0, 1\}$ that may predict a person's disease status.

Logistic regression and case-control studies

Suppose we sample people based on their disease status D ($D = 1$ is a **case**, $D = 0$ is a **control**).

We are interested in a binary marker $M \in \{0, 1\}$ that may predict a person's disease status.

The **prospective log odds**

$$\log \left[\frac{P(D = 1|M = m)}{P(D = 0|M = m)} \right] = \log \left[\frac{P(M = m|D = 1)P(D = 1)}{P(M = m|D = 0)P(D = 0)} \right]$$

measures how informative the marker is for the disease.

Logistic regression and case-control studies

Suppose we model $M|D$ using logistic regression, so

$$P(M = 1|D) = \frac{\exp(\alpha + \beta D)}{1 + \exp(\alpha + \beta D)}, \quad P(M = 0|D) = \frac{1}{1 + \exp(\alpha + \beta D)}.$$

Logistic regression and case-control studies

Suppose we model $M|D$ using logistic regression, so

$$P(M = 1|D) = \frac{\exp(\alpha + \beta D)}{1 + \exp(\alpha + \beta D)}, \quad P(M = 0|D) = \frac{1}{1 + \exp(\alpha + \beta D)}.$$

The prospective log odds can be written as

$$\log \left[\frac{\exp(M \cdot (\alpha + \beta)) / (1 + \exp(\alpha + \beta))}{\exp(M \cdot \alpha) / (1 + \exp(\alpha))} \cdot \frac{P(D = 1)}{P(D = 0)} \right]$$

Logistic regression and case-control studies

Suppose we model $M|D$ using logistic regression, so

$$P(M = 1|D) = \frac{\exp(\alpha + \beta D)}{1 + \exp(\alpha + \beta D)}, \quad P(M = 0|D) = \frac{1}{1 + \exp(\alpha + \beta D)}.$$

The prospective log odds can be written as

$$\log \left[\frac{\exp(M \cdot (\alpha + \beta)) / (1 + \exp(\alpha + \beta))}{\exp(M \cdot \alpha) / (1 + \exp(\alpha))} \cdot \frac{P(D = 1)}{P(D = 0)} \right]$$

which equals

$$\beta M + \log \left[\frac{1 + \exp(\alpha)}{1 + \exp(\alpha + \beta)} \cdot \frac{P(D = 1)}{P(D = 0)} \right].$$

Logistic regression and case-control studies

For prospective data and using logistic regression to model the prospective relationship $D|M$, the log odds would have the form

$$\theta + \beta M.$$

Logistic regression and case-control studies

For prospective data and using logistic regression to model the prospective relationship $D|M$, the log odds would have the form

$$\theta + \beta M.$$

The coefficient β when we use logistic regression to regress M on D using case-control data is the same coefficient (in the population sense) as we would obtain from regressing D on M in a prospective study.

Note that the intercepts are not the same in general

Estimation and inference for logistic regression

Assuming independent cases, the log-likelihood for logistic regression is

$$\begin{aligned} L(\beta | Y, X) &= \log \prod_i \frac{\exp(Y_i \cdot \beta^\top X_i)}{1 + \exp(\beta^\top X_i)} = \\ &= \sum_{i: Y_i=1} \beta^\top X_i - \sum_i \log(1 + \exp(\beta^\top X_i)). \end{aligned}$$

This likelihood is for the conditional distribution of Y given X .

As in linear regression, we do not model the marginal distribution of X .

Estimation and inference for logistic regression

The gradient of the log-likelihood function (the **score function**) is

$$G(\beta|Y, X) = \frac{\partial}{\partial \beta} L(\beta|Y, X) = \sum_{i: Y_i=1} X_i - \sum_i \frac{\exp(\beta^\top X_i)}{1 + \exp(\beta^\top X_i)} X_i.$$

Estimation and inference for logistic regression

The gradient of the log-likelihood function (the **score function**) is

$$G(\beta|Y, X) = \frac{\partial}{\partial \beta} L(\beta|Y, X) = \sum_{i: Y_i=1} X_i - \sum_i \frac{\exp(\beta^\top X_i)}{1 + \exp(\beta^\top X_i)} X_i.$$

The Hessian of the log-likelihood is

$$H(\beta|Y, X) = \frac{\partial^2}{\partial \beta \partial \beta^\top} L(\beta|Y, X) = - \sum_i \frac{\exp(\beta^\top X_i)}{(1 + \exp(\beta^\top X_i))^2} X_i X_i^\top.$$

The Hessian is strictly negative definite as long as the design matrix has independent columns.

Estimation and inference for logistic regression

The gradient of the log-likelihood function (the **score function**) is

$$G(\beta|Y, X) = \frac{\partial}{\partial \beta} L(\beta|Y, X) = \sum_{i: Y_i=1} X_i - \sum_i \frac{\exp(\beta^\top X_i)}{1 + \exp(\beta^\top X_i)} X_i.$$

The Hessian of the log-likelihood is

$$H(\beta|Y, X) = \frac{\partial^2}{\partial \beta \partial \beta^\top} L(\beta|Y, X) = - \sum_i \frac{\exp(\beta^\top X_i)}{(1 + \exp(\beta^\top X_i))^2} X_i X_i^\top.$$

The Hessian is strictly negative definite as long as the design matrix has independent columns.

Therefore $L(\beta|Y, X)$ is a concave function of β , so has a unique maximizer, and hence the MLE is unique.

Estimation and inference for logistic regression

From general theory about the MLE, the Fisher information

$$I(\beta) = -[EH(\beta|Y, X)|X]^{-1}$$

is the asymptotic sampling covariance matrix of the MLE $\hat{\beta}$.

Estimation and inference for logistic regression

From general theory about the MLE, the Fisher information

$$I(\beta) = -[EH(\beta|Y, X)|X]^{-1}$$

is the asymptotic sampling covariance matrix of the MLE $\hat{\beta}$.

Properties of MLE:

- ▶ $H(\beta|Y, X)$ does not depend on Y , therefore $I(\beta) = -H(\beta|X)^{-1}$.
- ▶ $\hat{\beta}$ is an MLE for a regular problem, it is consistent, asymptotically unbiased, and asymptotically normal if the model is correctly specified.

Components of GLM

- ▶ **Random Component:** Probability distribution for Y .
- ▶ **Systematic component:** Linear predictor:

$$\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

- ▶ **Link function:** Connects $E(Y) = \mu$ to the linear predictor

$$g(\mu) = \eta.$$

- ▶ **Variance:** A function that describes how the variance, $\text{var}(Y)$ depends on the mean

$$\text{var}(Y) = \varphi V(\mu).$$

where the dispersion parameter φ is a constant

Normal Regression

- ▶ $Y \sim N(\mu, \sigma^2),$
- ▶ $E(Y|X) = \mu,$
- ▶ $\mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p,$
- ▶ Identity link: $g(\mu) = \mu,$
- ▶ Variance function: $V(\mu) = 1.$

Logistic Regression

- ▶ $Y \sim \text{Bin}(n, p)$,
- ▶ We want to model the proportions Y/n ,
- ▶ $E(Y/n|X) = \mu = p$, $\text{var}(Y) = \frac{1}{n}p(1 - p)$,
- ▶ $\log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$,
- ▶ The logit link: $g(\mu) = \log \frac{\mu}{1-\mu}$,
- ▶ Variance function: $V(\mu) = \mu(1 - \mu)$.

Transformation vs. GLM

In some situations a response variable can be transformed to improve linearity and homogeneity of variance so that a general linear model can be applied. This approach has some drawbacks

Transformation vs. GLM

In some situations a response variable can be transformed to improve linearity and homogeneity of variance so that a general linear model can be applied. This approach has some drawbacks

- ▶ response variable has changed!
- ▶ transformation must simultaneously improve linearity and homogeneity of variance
- ▶ transformation may not be defined on the boundaries of the sample space

Transformation vs. GLM

For example, a common remedy for the variance increasing with the mean is to apply the log transform, e.g.

$$\log Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Which yields

$$E(\log Y) = \beta_0 + \beta_1 x_1$$

Transformation vs. GLM

For example, a common remedy for the variance increasing with the mean is to apply the log transform, e.g.

$$\log Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

Which yields

$$E(\log Y) = \beta_0 + \beta_1 x_1$$

This is a linear model for the mean of $\log Y$ which may not always be appropriate. E.g. if Y is income perhaps we are really interested in the mean income of population subgroups, in which case it would be better to model $E(Y)$ using a glm:

$$\log E(Y) = \beta_0 + \beta_1 x_1$$

with $V(\mu) = \mu$. This also avoids difficulties with $y = 0$

The distribution of Y : exponential family

Most of the commonly used statistical distributions, e.g. Normal, Binomial and Poisson, are members of the exponential family of distributions whose densities can be written in the form

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

where φ is the **dispersion** parameter and θ is the **canonical** parameter.

The distribution of Y : exponential family

Most of the commonly used statistical distributions, e.g. Normal, Binomial and Poisson, are members of the exponential family of distributions whose densities can be written in the form

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

where φ is the **dispersion** parameter and θ is the **canonical** parameter.

It can be shown that

$$E(Y) = b'(\theta) = \mu,$$

and

$$\text{var}(Y) = \varphi b''(\theta) = \varphi V(\mu).$$

Canonical Links

For a GLM where the response follows an exponential distribution we have

$$g(\mu) = g(b'(\theta)) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Canonical Links

For a GLM where the response follows an exponential distribution we have

$$g(\mu) = g(b'(\theta)) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

The **canonical** link is defined as

$$g = (b')^{-1},$$

so

$$g(\mu) = \theta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Canonical Links

For a GLM where the response follows an exponential distribution we have

$$g(\mu) = g(b'(\theta)) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

The **canonical** link is defined as

$$g = (b')^{-1},$$

so

$$g(\mu) = \theta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

Canonical links lead to desirable statistical properties of the GLM hence tend to be used by default. However there is no a priori reason why the systematic effects in the model should be additive on the scale given by this link.

Estimation of Model Parameters

A single algorithm can be used to estimate the parameters of an exponential family GLM using maximum likelihood.

The log-likelihood for the sample $Y_1, \dots, Y_n$ is

$$L = \sum_{i=1}^n \frac{Y_i \theta_i - b(\theta_i)}{\theta_i} + c(Y_i, \varphi_i)$$

Estimation of Model Parameters

A single algorithm can be used to estimate the parameters of an exponential family GLM using maximum likelihood.

The log-likelihood for the sample $Y_1, \dots, Y_n$ is

$$L = \sum_{i=1}^n \frac{Y_i \theta_i - b(\theta_i)}{\theta_i} + c(Y_i, \varphi_i)$$

The maximum likelihood estimates are obtained by solving the score equations

$$s(\beta_j) = \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^n \frac{Y_i - \mu_i}{\varphi_i V(\mu_i)} \times \frac{x_{ij}}{g'(\mu_i)} = 0$$

Estimation of Model Parameters

We assume that

$$\varphi_i = \frac{\varphi}{a_i}$$

where φ is a single dispersion parameter and a_i are known prior weights; for example binomial proportions with known index n_i have $\varphi = 1$ and $a_i = n_i$.

Estimation of Model Parameters

We assume that

$$\varphi_i = \frac{\varphi}{a_i}$$

where φ is a single dispersion parameter and a_i are known prior weights; for example binomial proportions with known index n_i have $\varphi = 1$ and $a_i = n_i$.

The estimating equations are then

$$s(\beta_j) = \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^n \frac{a_i(Y_i - \mu_i)}{V(\mu_i)} \times \frac{x_{ij}}{g'(\mu_i)} = 0$$

which does not depend on φ (which may be unknown).

Fisher Scoring

A general method of solving score equations is the iterative algorithm **Fishers Method of Scoring** (derived from a Taylor's expansion of $s(\beta)$)

In the r -th iteration, the new estimate $\beta^{(r+1)}$ is obtained from the previous estimate $\beta^{(r)}$ by

$$\beta^{(r+1)} = \beta^{(r)} + s(\beta^{(r)})E \left(H(\beta^{(r)}) \right)^{-1},$$

where H is the Hessian of the log-likelihood.

It turns out that the updates can be written as

$$\beta^{(r+1)} = (X^T W^{(r)} X)^{-1} X^T W^{(r)} z^{(r)},$$

i.e. the score equations for a weighted least squares regression of $z^{(r)}$ on X with weights $W^{(r)} = \text{diag}(w_i)$, where

$$z_i^{(r)} = \eta_i^{(r)} + (Y_i - \mu_i^{(r)}) g'(\mu_i^{(r)})$$

and

$$w_i^{(r)} = \frac{a_i}{V(\mu_i^{(r)}) \left(g'(\mu_i^{(r)}) \right)^2}$$

IRWLS

Hence the estimates can be found using an Iteratively (Re-)Weighted Least Squares algorithm:

1. Start with initial estimates $\mu_i^{(r)}$
2. Calculate working responses $z_i^{(r)}$ and working weights $w_i^{(r)}$
3. Calculate $\beta^{(r+1)}$ by weighted least squares
4. Repeat 2 and 3 till convergence

For models with the canonical link, this is simply the Newton-Raphson method.

Standard Errors

The estimates $\hat{\beta}$ have the usual properties of maximum likelihood estimators. In particular, $\hat{\beta}$ is asymptotically

$$N(\beta, i(\beta)^{-1}),$$

where

$$i(\beta) = \varphi^{-1} X^{\top} W X$$

Standard Errors

The estimates $\hat{\beta}$ have the usual properties of maximum likelihood estimators. In particular, $\hat{\beta}$ is asymptotically

$$N(\beta, i(\beta)^{-1}),$$

where

$$i(\beta) = \varphi^{-1} X^T W X$$

Standard errors for the β_j may therefore be calculated as the square roots of the diagonal elements of

$$\text{cov}(\hat{\beta}) = \varphi (X^T W X)^{-1}$$

in which $(X^T W X)^{-1}$ is a by-product of the final IRWLS iteration.

If φ is unknown, an estimate is required.

Dispersion

There are practical difficulties in estimating the dispersion φ by maximum likelihood.

Therefore it is usually estimated by method of moments. If β was known, an unbiased estimate of $\varphi = a_i \text{var}(Y_i)/V(\mu_i)$ would be

$$\frac{1}{n} \sum_{i=1}^n \frac{a_i (Y_i - \mu_i)^2}{V(\mu_i)},$$

and since β is to be estimated we end with

$$\hat{\varphi} = \frac{1}{n-p} \sum_{i=1}^n \frac{a_i (Y_i - \mu_i)^2}{V(\mu_i)},$$

Examples of GLM's

Poisson regression: the distribution of $Y|X$ follows a Poisson distribution, with the mean response related to the covariates via

$$\log E(Y|X) = \beta^T X.$$

It follows that $\log \text{var}(Y|X) = \beta^T X$ as well.

Examples of GLM's

Poisson regression: the distribution of $Y|X$ follows a Poisson distribution, with the mean response related to the covariates via

$$\log E(Y|X) = \beta^T X.$$

It follows that $\log \text{var}(Y|X) = \beta^T X$ as well.

Negative binomial regression: the probability mass function for the dependent variable Y is

$$P(Y_i = y|X) = \frac{\Gamma(y + 1/\alpha)}{\Gamma(y + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha\mu_i} \right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i} \right)^y.$$

The mean of this distribution is μ_i and the variance is $\mu_i + \alpha\mu_i^2$. If $\alpha = 0$ we get the same mean/variance relationship as the Poisson model. As α increases, we get increasingly more overdispersion.

Overdispersion

Under the Poisson model, $\text{var}(Y|X) = E(Y|X)$. A Poisson model results from using the Poisson GLM with the scale parameter ϕ fixed at 1.

Overdispersion

Under the Poisson model, $\text{var}(Y|X) = E(Y|X)$. A Poisson model results from using the Poisson GLM with the scale parameter ϕ fixed at 1.

The **quasi-Poisson** model is the Poisson model with a scale parameter that may be any non-negative value. Under the quasi-Poisson model, $\text{var}(Y|X) \propto E(Y|X)$.

The negative binomial GLM allows the variance to be non-proportional to the mean.

Overdispersion

Under the Poisson model, $\text{var}(Y|X) = E(Y|X)$. A Poisson model results from using the Poisson GLM with the scale parameter ϕ fixed at 1.

The **quasi-Poisson** model is the Poisson model with a scale parameter that may be any non-negative value. Under the quasi-Poisson model, $\text{var}(Y|X) \propto E(Y|X)$.

The negative binomial GLM allows the variance to be non-proportional to the mean.

Any situation in which $\text{var}(Y|X) > E(Y|X)$ is called **overdispersion**. Overdispersion is often seen in practice.

Wald Tests

The Wald test follows immediately from the fact that the large sample distribution of the maximum likelihood estimator $\hat{\beta}$ is multivariate normal:

$$\hat{\beta} \sim N_p(\beta, \varphi (X^T W X)^{-1})$$

Tests for subsets of β are based on the corresponding marginal normal distributions.

Wald Tests

The Wald test follows immediately from the fact that the large sample distribution of the maximum likelihood estimator $\hat{\beta}$ is multivariate normal:

$$\hat{\beta} \sim N_p(\beta, \varphi (X^T W X)^{-1})$$

Tests for subsets of β are based on the corresponding marginal normal distributions.

Specifically, we test

$$H_0 : \beta_j = 0 \quad \text{versus} \quad H_1 : \beta_j \neq 0$$

using the test statistic

$$z_j = \frac{\hat{\beta}_j}{\sqrt{\varphi (X^T W X)^{-1}_{jj}}}$$

which is asymptotically $N(0, 1)$ under H_0 .

Deviance

The **deviance** of a model is defined as

$$D = 2\varphi(L_{sat} - L_{mod})$$

where L_{mod} is the log-likelihood of the fitted model and L_{sat} is the log-likelihood of the saturated model.

In the saturated model, the number of parameters is equal to the number of observations, so $\hat{Y} = Y$.

For linear regression with Normal data, the deviance is equal to the residual sum of squares.

Model comparison for GLM's

In the Gaussian case, $D/\varphi = L \sim \chi^2_{n-p-1}$.

When φ is unknown, we can turn this around to produce an estimate of the scale parameter

$$\hat{\varphi} = \frac{D}{n - p - 1}.$$

This is an unbiased estimate in the Gaussian case, but is useful for any GLM.

Model comparison for GLM's

Now suppose we want to compare two nested generalized linear models with deviances $D_1 < D_2$. Let $p_1 > p_2$ be the number of covariates in each model. The likelihood ratio test statistic is

$$L_2 - L_1 = \frac{D_2 - D_1}{\varphi}$$

which asymptotically has a $\chi^2_{p_1 - p_2}$ distribution.

If φ is unknown, we can estimate it as described above (using the larger of the two models).

The “plug-in” likelihood ratio statistic $(D_2 - D_1)/\hat{\varphi}$ is still asymptotically $\chi^2_{p_1 - p_2}$, as long as $\hat{\varphi}$ is consistent.

The finite sample distribution may be better approximated using

$$\frac{D_2 - D_1}{\hat{\varphi}(p_1 - p_2)} \approx F_{p_1 - p_2, n - p_1},$$

which is exact in the Gaussian case.