

ECON 613: Applied Econometrics

Methods

Overview: Linear Models

- ▶ Study the relationship between an outcome variable y and a set of regressors x .
 - ▶ Conditional Prediction.
 - ▶ Causal inference.
 - ▶ Example: propensity to consume.
- ▶ Loss function approach

$$L(e) = L(y - \hat{y})$$

where $\hat{y} = E(y \mid x)$ is a predictor of y , and the error $e = y - \hat{y}$

Squared Loss Function

- ▶ Squared error loss: $L(e) = e^2$
- ▶ Optimization problem

$$\min_{\beta} \sum_i^N (y_i - f(x_i, \beta))^2$$

Linear Prediction

- ▶ $E[y \mid x] = x'\beta$
- ▶ OLS

$$y = x\beta + e$$

- ▶ Derivation

$$\begin{aligned} L(\beta) &= (y - x\beta)'(y - x\beta) \\ &= y'y - 2y'x\beta + \beta'x'x\beta \end{aligned}$$

Then

$$\frac{\partial L(\beta)}{\partial \beta} = -2x'y + 2x'x\beta = 0$$

- ▶ Formula

$$\hat{\beta} = (x'x)^{-1}x'y$$

Other loss functions?

- ▶ Median regression, also known as least-absolute-deviations (LAD) regression, minimizes

$$\sum_i |e_i|$$

- ▶ No close form solution.

Properties

see 4.4.4 and 4.4.5.

Properties of an estimator

- ▶ Unbiasedness: $E(\hat{\theta}) = \theta$.
- ▶ Consistency: $\text{plim}\hat{\theta}_n = \theta$.
- ▶ Efficiency: Reach Cramer-Rao lower bound asymptotically.

Application: More guns, less crime?

- ▶ Using cross sectional time series data for US counties from 1977 to 1992, we find that allowing citizens to carry concealed weapons deters violent crimes, without increasing accidental deaths.
- ▶ Thoughts?

A dataset: Guns in the AER package

- ▶ state: factor indicating state.
- ▶ year: factor indicating year.
- ▶ violent: violent crime rate (incidents per 100,000).
- ▶ murder: murder rate (incidents per 100,000).
- ▶ robbery: robbery rate (incidents per 100,000).
- ▶ prisoners: incarceration rate in the state in the previous year.
- ▶ afam: percent of state population that is African-American.
- ▶ cauc: percent of state population that is Caucasian.
- ▶ male: percent of state population that is male.
- ▶ population: state population.
- ▶ income: real per capita personal income in the state (\$).
- ▶ density: population per square mile of land area.
- ▶ law: factor. Does the state have a shall carry law in effect in that year?

Table: Statistical models

	Model 1	Model 2	Model 3	Model 4
(Intercept)	6.13*** (0.02)	2.98*** (0.54)	4.04*** (0.39)	3.97*** (0.47)
lawyes	-0.44*** (0.04)	-0.37*** (0.03)	-0.05* (0.02)	-0.03 (0.02)
prisoners		0.00*** (0.00)	-0.00 (0.00)	0.00 (0.00)
density		0.03* (0.01)	-0.17* (0.09)	-0.09 (0.08)
income		0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
population		0.04*** (0.00)	0.01 (0.01)	-0.00 (0.01)
afam		0.08*** (0.02)	0.10*** (0.02)	0.03 (0.02)
cauc		0.03*** (0.01)	0.04*** (0.01)	0.01 (0.01)
Year and state FE				X
R ²	0.09	0.56	0.94	0.96
Adj. R ²	0.09	0.56	0.94	0.95
Num. obs.	1173	1173	1173	1173
RMSE	0.62	0.43	0.16	0.14

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Model Selection

- ▶ *R – squared*
- ▶ Model selection criterias: AIC, BIC
- ▶ Endogeneity concerns.

Principle of a monte-carlo study

- ▶ Generate S independent data sets under the conditions of interest
- ▶ Compute the numerical value of the estimator/test statistic $T(\text{data})$ for each data set T_1, \dots, T_S
- ▶ If S is large enough, summary statistics across T_1, \dots, T_S .

Monte-carlo

- ▶ Simulate $xvec$ drawing from a normal distribution
- ▶ Set $a=2$, and $b=0.1$, and construct $yvec = a + bxvec + \epsilon$ with $\epsilon \sim Normal(0, 1)$.

Potential problems: sample size

Table: Statistical models

	$n = 10$	$n = 100$	$n = 1000$	$n = 10000$	$n = 100000$
(Intercept)	1.69*** (0.28)	1.90*** (0.10)	2.02*** (0.03)	1.99*** (0.01)	2.00*** (0.00)
xvec	1.01* (0.39)	-0.05 (0.11)	0.04 (0.03)	0.11*** (0.01)	0.10*** (0.00)
R ²	0.45	0.00	0.00	0.01	0.01
Adj. R ²	0.38	-0.01	0.00	0.01	0.01
Num. obs.	10	100	1000	10000	100000
RMSE	0.85	0.96	1.01	1.00	1.00

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Miss-specification (1)

- ▶ Simulate $xvec$ drawing from a normal distribution
- ▶ Set $a=2$, and $b=0.1$, and construct $yvec = a + bxvec + \epsilon$ with $\epsilon \text{ uniformly}(0, 5)$.
- ▶ Thoughts?

Error terms misspecification

Table: Statistical models

	$n = 10$	$n = 100$	$n = 1000$	$n = 10000$	$n = 10000$
(Intercept)	5.06*** (0.41)	4.37*** (0.15)	4.50*** (0.05)	4.48*** (0.01)	4.50*** (0.00)
xvec	0.40 (0.45)	0.26 (0.17)	0.11* (0.05)	0.10*** (0.01)	0.10*** (0.00)
R ²	0.09	0.02	0.01	0.00	0.00
Adj. R ²	-0.03	0.01	0.00	0.00	0.00
Num. obs.	10	100	1000	10000	100000
RMSE	1.30	1.53	1.45	1.44	1.45

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Miss-specification (2)

- ▶ Simulate $xvec$ drawing from a normal distribution
- ▶ Set $a=2$, and $b=0.1$, and construct
 $yvec = a + b \exp(xvec) + \epsilon$ with $\epsilon \sim Normal(0, 1)$.

Functional form misspecification

Table: Statistical models

	$n = 10$	$n = 100$	$n = 1000$	$n = 10000$	$n = 10000$
(Intercept)	5.21*** (0.41)	4.52*** (0.15)	4.67*** (0.05)	4.64*** (0.01)	4.66*** (0.00)
xvec	0.48 (0.45)	0.31 (0.17)	0.18*** (0.05)	0.17*** (0.01)	0.17*** (0.00)
R ²	0.12	0.03	0.02	0.01	0.01
Adj. R ²	0.02	0.03	0.01	0.01	0.01
Num. obs.	10	100	1000	10000	100000
RMSE	1.29	1.52	1.46	1.45	1.45

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Maximum Likelihood

GMM

Numerical Optimization

Inference

Introduction to MLE

Consider a parametric model in which the joint distribution of $Y = (Y_1, \dots, Y_n)$ has a density $\ell(y, \theta)$ with respect to a measure μ . Then consider $P_\theta = \ell(y, \theta)\mu$ where $\theta \in \Theta \in \mathbb{R}^p$. Once $y = (y_1, \dots, y_n)$ is observed, the maximum likelihood method consists of estimating the parameter θ a value $\hat{\theta}(y)$ that maximizes the likelihood function $\theta \rightarrow \ell(y, \theta)$. Formally, a maximum likelihood estimator of θ is a solution to the maximization problem

$$\max_{\theta} \ell(Y; \theta)$$

or

$$\max_{\theta} \log(\ell(Y; \theta))$$

Feasible examples: Poisson distribution

Consider a dependent variable that takes only non negative integer values $0, 1, 2, \dots$, and one assumes that the dependent variable follows a Poisson distribution, and we wish to estimate the Poisson parameter.

- ▶ Given $y_i \sim f(\lambda, y_i) = \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}$
- ▶ Likelihood $\mathcal{L}(y; \lambda) = \prod_{i=1}^N \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!} = \frac{\exp(-N\lambda)\lambda^{\sum_{i=1}^N y_i}}{\prod_{i=1}^N y_i!}$
- ▶ Log likelihood
 $\log \mathcal{L}(y; \lambda) = -N\lambda + \sum_{i=1}^N y_i \log(\lambda) - \sum_{i=1}^N \log(y_i!)$
- ▶ Estimate

$$\frac{\partial \log \mathcal{L}(y; \lambda)}{\partial \lambda} = 0 \implies \hat{\lambda} = \frac{\sum_{i=1}^N y_i}{N}$$

Feasible examples: Least Squares

- ▶ Normality assumption $e \sim \mathbb{N}(0, \sigma^2)$, then $y \sim \mathbb{N}(x\beta, \sigma^2)$.
- ▶ Likelihood $L(\beta) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp(-0.5\sigma^{-2}(y - x\beta)'(y - x\beta))$
- ▶ log likelihood $\log L(\beta) = -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta)$
- ▶ $\beta = (x'x)^{-1}x'y$

Some difficulties

- ▶ Non-uniqueness of the Likelihood Function
- ▶ Non-existence of a solution to the Maximization Problem
- ▶ Multiple Solutions to the Maximization Problem

Asymptotic Properties (1): Convergence

Definition

Under a set of regularity conditions, there exists a sequence of maximum likelihood estimators converging almost surely to the true parameter value θ_0

- ▶ The variables $Y_i, i = 1, 2, \dots$ are independent and identically distributed with density $f(y; \theta), \theta \in \Theta \in \mathbb{R}^p$
- ▶ The parameter space Θ is compact.
- ▶ The log likelihood function $\mathcal{L}(y, \theta)$ is continuous in θ and is a measurable function of y .
- ▶ The log-likelihood function is such that $(1/n)\mathcal{L}_n(y, \theta)$ converges surely to $E_{\theta_0} \log(f(Y_i; \theta))$ uniformly in $\theta \in \Theta$. $E_{\theta_0} \log(f(Y_i; \theta))$ exists.

Asymptotic Properties (2): Asymptotic Normality

- ▶ The log likelihood function $\mathcal{L}_n(\theta)$ is twice continuously differentiable in an open neighborhood of θ_0
- ▶ The matrix (Fisher Information Matrix)

$$\mathcal{I}_1(\theta_0) = E_{\theta_0} \left(- \frac{\partial^2 \log f(Y_1; \theta_0)}{\partial \theta \partial \theta'} \right)$$

Definition

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow \mathbb{N}(0, \mathcal{I}_1(\theta_0)^{-1}).$$

Concentrated Likelihood Function

Definition

Let the parameter set $\theta = (\alpha, \beta)$. The solutions $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ to the maximization problem $\max_{\alpha, \beta} \log \mathcal{L}(y; \alpha, \beta)$ can be obtained via the following two-step procedure:

- a) Maximize the log-likelihood function with respect α given β . The maximum value is attained for values of α in a set $A(\beta)$ depending on the parameter β . Thus, if $\alpha \in A(\beta)$, the log-likelihood value is

$$\log \mathcal{L}_c(y; \beta) = \max_{\alpha} \log \mathcal{L}(y; \alpha, \beta)$$

The mapping $\log \mathcal{L}_c$ is called the concentrated (in α) log likelihood function.

- b) In a second step, maximize the concentrated log-likelihood function with respect to β .

Application

Consider the likelihood

$$\mathcal{L}(y, \beta, \sigma) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - x\beta)'(y - x\beta)$$

► First step

$$\frac{\partial \mathcal{L}(y; \beta, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} (y - x\beta)'(y - x\beta) = 0$$

Then

$$\sigma^2(\beta) = \frac{1}{n} (y - x\beta)'(y - x\beta)$$

► Substituting $\sigma^2(\beta)$ into the likelihood

$$\mathcal{L}_c(y, \beta, \sigma) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \frac{1}{n} (y - x\beta)'(y - x\beta) - \frac{n}{2}$$

Hypothesis Testing

Three procedures to do tests

Likelihood Ratio

- ▶ The likelihood ratio statistic is

$$LR = 2(\ell(\theta) - \ell(\tilde{\theta}))$$

where $\hat{\theta}$ and $\tilde{\theta}$ are the restricted and unrestricted maximum likelihood estimates of θ .

- ▶ Wilk's theorem shows that

$$LR \sim \chi^2(r)$$

where r is the number of restrictions.

Additional Tests

- ▶ Wald Test
- ▶ LM test

We will see in GMM.

In practice

- ▶ The regularity conditions are strong.
- ▶ What happens if we weaken them?

Problem 1

Number of parameters increases with the number of observations

- ▶ Convergence holds
- ▶ Estimates may be biased

Problem 2

True parameter value θ_0 does not belong to Θ : The model is misspecified

- ▶ Convergence holds to a parameter that is not the true parameter.

Problem 3

Correlated Observations

- ▶ Convergence does not hold.

Problem 4

Discontinuity of the likelihood function

- ▶ Numerical problems.

Problem 5

Known parameter space

- ▶ Constrained Optimization

Maximum Likelihood

GMM

Numerical Optimization

Inference

Method of Moments

- ▶ Orthogonality condition in Linear Models

$$E(x(y' - x)) = 0 \quad (1)$$

- ▶ Moment Condition

$$\frac{1}{N} \sum_i x_i (y_i - x_i' \beta) \quad (2)$$

- ▶ Moment Estimator

$$\hat{\beta}_{\text{MM}} = \left(\sum_i x_i x_i' \right)^{-1} \left(\sum_i x_i y_i \right) \quad (3)$$

Nonlinear Model

- ▶ Consider

$$Y_i = g(X_i, b_0) + u_i$$

- ▶ Orthogonality Condition

$$E[X'(y - g(X, b_0))] = 0$$

- ▶ Moments condition

$$E_0 h(Y, X, a_0) = 0$$

- ▶ The function h is H -dimensional and the parameter a is of size K .

Formal Idea

Definition

The basic idea of generalized method of moments is to choose a value for a such that the sample mean is closest to zero.

$$\frac{1}{n} \sum_{i=1}^n h(Y_i, X_i, a)$$

Formal Definition

Definition

Let \mathbb{S}_n be an $(H \times H)$ symmetric positive definite matrix that may depend on the observations. The generalized method of moments (GMM) estimator associated with \mathbb{S}_n is a solution $\tilde{a}_n(\mathbb{S}_n)$ to the problem

$$\min_a \left[\sum_{i=1}^n h(Y_i, X_i, a) \right]' \mathbb{S}_n \left[\sum_{i=1}^n h(Y_i, X_i, a) \right]$$

Assumptions

- H1 The variables (Y_i, X_i) are independent and identically distributed.
- H2 The expectation $E_0 h(Y, X, a)$ exists and is zero when a is equal to the true value a_0 of the parameter of interest.
- H3 The matrix S_n converges almost surely to a nonrandom matrix S_0
- H4 The parameter a_0 is identified from the equality constraints, i.e. $E_0 h(Y, X, a)' S_0 E_0 h(Y, X, a) = 0$
- H5 The parameter value a_0 is known to belong to a compact set \mathcal{A}
- H6 The quantity $(1/n) \sum_{i=1}^n h(Y_i, X_i, a)$ converges almost surely and uniformly in a to $E_0 h(Y, X, a)$
- H7 The function $h(Y, X, a)$ is continuous in a
- H8 The matrix $\left[E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 \left[E_0 \frac{h(Y, X, a)}{\partial a'} \right]$ is nonsingular, which implies $H \geq K$.

Asymptotic Normality

Under the assumptions, we have

$$\sqrt{n}(\tilde{a}_n(S_n) - a_0) \sim \mathbb{N}(0, \Sigma(S_0))$$

where

$$\begin{aligned} \Sigma(S_0) = & \left(\left[E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 \left[E_0 \frac{h(Y, X, a)}{\partial a'} \right] \right)^{-1} \\ & \left(\left[E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 V_0(h(Y, X, a_0)) S_0 \left[E_0 \frac{h(Y, X, a)}{\partial a'} \right] \right)^{-1} \\ & \left(\left[E_0 \frac{h(Y, X, a)}{\partial a} \right]' S_0 \left[E_0 \frac{h(Y, X, a)}{\partial a'} \right] \right)^{-1} \end{aligned}$$

Optimal GMM

- ▶ \mathbb{S}_0 is not known.
- ▶ Two-step procedure
 - ▶ Estimate

$$\min_a \left[\sum_{i=1}^n h(Y_i, X_i, a) \right]' I \left[\sum_{i=1}^n h(Y_i, X_i, a) \right]$$

where I is the identity matrix, and recover \hat{a} .

- ▶ Matrix of variance/covariance

$$\hat{\mathbb{S}} = \frac{1}{N} \sum_{i=1}^n h(Y_i, X_i, \hat{a}) h(Y_i, X_i, \hat{a})'$$

Relationship to IV.

- ▶ Nonlinear 2SLS is a very good application of GMM.

Inference

- ▶ Over identification test see iv section.

Applications

- ▶ Matrix of variance/covariance in practice
- ▶ Indirect Inference
- ▶ Simulated method of moments

Maximum Likelihood

GMM

Numerical Optimization

Inference

Numerical Optimization

Most maximum likelihood estimates require numerical optimization.

Primer on optimization

Definition

$$\min_x f(x)$$

- ▶ $x \in \mathbb{R}^n$
- ▶ f is a smooth function.

Existence: Weierstrass theorem

A point or a vector x^* is a global minimizer if $f(x^*) \leq f(x) \forall x$.

Maximization Vs Minimization

Let $-f$ denote the function whose value at any value at any x is $-f(x)$. Then,

1. x is the maximum of f if and only if x is a minimum of $-f$
2. z is a minimum of f if and only if z is a maximum of $-f$

Necessary conditions

1. If x^* is local minimizer and f is continuously differentiable in an open neighborhood of x^* , then $\nabla f(x^*) = 0$
2. If x^* is local minimizer and $\nabla^2 f$ exists and is continuous in an open neighborhood of x^* , then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) = 0$

Likelihood setup

The likelihood function is defined by:

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n f(X_i; \theta)$$

The necessary conditions for optimization yield the regularity conditions

Unfeasible example:

Any nonlinear model

Numerical optimization

- ▶ Local optimization: the best minimum/maximum in a vicinity
- usually defined by a convergence criteria.
- ▶ Global optimization: Best of all local minimas/maximas.

Numerical optimization - Local Optimization

Overview

1. **Line Search:** Starting from an initial value, choose a direction and search along this direction to find a new iterate
2. **Trust region:** Use previous estimates of the objective function, to construct a **synthetic** or **model** function whose behavior near the current point is similar to the objective function, and search only over a region, *trust region*, with the underlying idea that the model function is a good approximate over the trust region.

Line Search

Idea:

$$x_{k+1} = x_k + \alpha_k d_k$$

where d_k is a direction to be evaluated, and α_k a scaling parameter.

The variants of numerical optimization

1. Steepest descent: $d_k = -\nabla f(x_k)$
2. Newton direction: $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$
3. Quasi-Newton direction: $d_k = -(B(x_k))^{-1} \nabla f(x_k)$
4. Derivative free.

Properties

- ▶ Robustness: Perform well for various problems and starting values.
- ▶ Efficiency:
- ▶ Accuracy: Identify a solution with precision, not sensitive to starting values

One parameter optimization

- ▶ Bisection
- ▶ Secant Method

Codes

```
bisection <- function(f, a, b, n , tol ) {  
  # Check the signs of the function.  
  if (!(f(a) < 0) && (f(b) > 0)) {  
    stop()} else if ((f(a) > 0) && (f(b) < 0)) {  
    stop()}  
  for (i in 1:n) {  
    c <- (a + b) / 2 # Calculate midpoint  
    # If the function equals 0 at the midpoint  
    if ((f(c) == 0) || ((b - a) / 2) < tol) {  
      return(c) }  
    # If another iteration is required,  
    # check the signs of the function  
    ifelse(sign(f(c)) == sign(f(a)),  
           a <- c,  
           b <- c)  
  }  
  # If the max number of iterations is reached  
  print('Too many iterations')
```

Recover the scaling parameter

- Solve the function $\phi(\alpha) = f(x_k + \alpha d_k)$

Quasi-Newton methods

How to approximate the hessian such that:

- ▶ Reduce the computation time (Use only gradient instead of hessian)
- ▶ Increase convergence rate

Conjugate Gradient- FR

- ▶ Given x_0
- ▶ Evaluate $f_0 = f(x_0)$, $\nabla f_0 = \nabla f(x_0)$
- ▶ Set $d_0 = -\nabla f_0$, $k = 0$
- ▶ **While** $\nabla f_k \neq 0$
 - ▶ Set $x_{k+1} = x_k + \alpha_k d_k$
 - ▶ Evaluate ∇f_{k+1} , then:
 - ▶ $\beta_{k+1}^{FR} = \frac{\nabla f'_{k+1} \nabla f_{k+1}}{\nabla f'_k \nabla f_k}$
 - ▶ $d_{k+1} = -\nabla f_{k+1} + \beta_{k+1}^{FR} d_k$
 - ▶ $k = k + 1$
- ▶ **end(while)**

BFGS

- ▶ Given x_0
- ▶ Evaluate $f_0 = f(x_0)$, $\nabla f_0 = \nabla f(x_0)$, $H_0 = I$
- ▶ Set $k = 0$
- ▶ **While** $\|\nabla f_k\| > \epsilon$
 - ▶ Compute direction $d_k = -H_k \nabla f_k$
 - ▶ Set $x_{k+1} = x_k + \alpha_k d_k$
 - ▶ Evaluate ∇f_{k+1} , then:
 - ▶ set $s_k = x_{k+1} - x_k$, $y_k = \nabla f_{k+1} - \nabla f_k$ and $\rho_k = \frac{1}{y_k' s_k}$
 - ▶ Update $H_{k+1} = (I - \rho_k s_k y_k') H_k (I - \rho_k y_k s_k') + \rho_k s_k s_k'$
- ▶ **end(while)**

Problems

- ▶ non differentiable functions
- ▶ disconnected and non-convex feasible space
- ▶ discrete feasible space
- ▶ large dimensionality
- ▶ multiple local minimas

Derivative Free

- ▶ Nelder-Mead
- ▶ Simulated annealing
- ▶ Divided Rectangles Method
- ▶ Genetic Algorithms
- ▶ Particle Optimization

Maximum Likelihood

GMM

Numerical Optimization

Inference

How to get standard error

- ▶ Fisher Information Matrix
- ▶ Sandwich formula

Introduction to Bootstrap

- ▶ Inference for small samples basically...
- ▶ Inference for unknown distribution...
- ▶ Computer intensive resampling method...
 - ▶ Using data to generate new data

Applications: Inference in estimation

- ▶ Bootstrap samples
- ▶ Parallel implementation

Applications: Inference post estimation

- ▶ Bootstrap samples
- ▶ Marginal effects

Applications: Testing

- ▶ Bootstrap samples
- ▶ Approach to testing

Applications: Choosing the number of R

- ▶ Objectives and Constraints
- ▶ $R=49,99,199,499,999,9999....$

Delta Method

- ▶ Consider $X \sim \mathbb{N}(\mu, \sigma^2)$, and assume you are interested in $E(g(X))$ and $Var(g(X))$
- ▶ Approximation

$$g(x) = g(\mu) + g'(\mu)(x - \mu) \quad (4)$$

and then

$$E(g(x)) \approx g(E(x)) \quad (5)$$

$$Var(g(x)) \approx g'(E(x))^2 Var(x) \quad (6)$$

- ▶ Usage?