Mathematics 50

Probability and Statistical Inference

Winter 1999

Methods of Statistical Estimation

Week 8

February 22–26

1. Parameter estimation for bivariate distribution

We have a couple of random variables (X, Y) with a bivariate distribution $F(x, y)$.

We have n iid pairs drawn from $F : (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$.

Means and variances are unknown.

We can estimate marginal means and variances (biased version) as

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \\ \widehat{\sigma}_x^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2, \quad \widehat{\sigma}_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2. \end{aligned}$$

How to estimate the covariance? Recall

$$cov(X, Y) = E(X - \mu_X)(Y - \mu_Y),$$

so that the empirical version (estimator, or sample covariance) is

$$\widehat{\sigma}_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$$

1.1. Estimation of correlation coefficient (c.c.)

C.c measures how two RVs are stochastically (linearly) dependent, or *correlated,*

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

where

$$\sigma_{XY} = cov(X, Y)$$

Recall basic properties of c.c:

1. $|\rho| \leq 1$
2. $\rho = 0$ - uncorrelated
3. $\rho = 1$ positive linear functional relationship

4. $\rho = -1$ negative linear functional relationship

If the data are

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

sample correlation coefficient becomes

$$r = \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x \widehat{\sigma}_y}$$

Combining with previous estimators for $\widehat{\sigma}_{xy}$, $\widehat{\sigma}_x$ and $\widehat{\sigma}_y$ we come to an estimator for ρ :

$$r = \frac{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2}}.$$

r behaves like its theoretical counterpart:

$|r| \leq 1$

$r = \pm 1$ then $y_i = a + bx_i$

r does not change under any linear transformations of x or y.

1.2. Estimation of conditional mean (regression)

Recall, if (X, Y) have bivariate normal distribution then the conditional distribution of Y given X also has a normal distribution,

$$Y|(X = x) \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X), \sigma_{Y|X}^2\right).$$

The conditional mean is

$$E(Y|X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

How to estimate this function? Substitute the sample version (estimate) for the true unknown quantities.

Hence we come to empirical conditional mean (*regression*)

$$\begin{aligned} & \widehat{\mu}_y + r\frac{\widehat{\sigma}_y}{\widehat{\sigma}_x}(x - \widehat{\mu}_x) = \widehat{\mu}_y + \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x\widehat{\sigma}_y}\frac{\widehat{\sigma}_Y}{\widehat{\sigma}_X}(x - \widehat{\mu}_x) \\ = \quad & \widehat{\mu}_y + \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}(x - \widehat{\mu}) = \left(\widehat{\mu}_y - \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}\widehat{\mu}_x\right) + \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}x. \end{aligned}$$

Hence the slope of regression is

$$\frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}$$

and the intercept

$$\widehat{\mu}_y - \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}\widehat{\mu}_x.$$

Therefore, to estimate regression (or conditional mean) we estimate the slope as

$$b = \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2}$$

the intercept as

$$a = \widehat{\mu}_y - b\widehat{\mu}_x.$$

The regression line is a linear function of x :

$$a + bx.$$

The predicted values of Y given $X = x_i$ is

$$\widehat{y}_i = a + bx_i$$

2. STATA

is one of the most popular statistical packages. This package is easy to use and get started, and it is available at Dartmouth (on Public). I assume you installed STATA on you computer; instructions on the installation procedure have been sent out a week ago.

The most useful command for the first-time user is `exit.`

Commands `help` and `lookup` are for help. For example, if you want to get a help on 'save' you type

`help save` and press <Enter>.

To continue scrolling press <SpaceBar>. To quit help press `q.`

To input new data use `input` var1 var 2 var3, etc. For example, if you want type in grades for homework1 and homework2 you issue the following command

`input hw1 hw2`

Then you type in the data pressing <Enter> at the end of the line.

To end type `end.`

In STATA you can use missing observations. If a case is missing just type in period '.' Period means the case is missing (not recorded).

To see the data use `list.` Use <SpaceBar> or <q> to control scrolling.

To correct the data use `replace` varname `in` case. For example, if in the variable hmw2 in the 5-th case it should be 37 use

`replace hmw2 37 in 5`

To create a new variable as a function of existing variables use command

`generate.`

(You can use only first 3 letters.) For example, if var1 is an existing variable and you want to generate a squared variable you issue

`gen var2=var1^2`

You can use the following arithmetic operators: +,-,*, /,^ (power). Also the following mathematical functions are available:

`exp, ln, sqrt, abs`, etc.

Try `help functions` to see the list of what is available. For example, try this `gen var3=ln(var1^2+var1+1)`

The rule in STATA is: if you generate a new variable the length of that variable is equal to the length of old variables (it implies that all variable in the current data set have the same length).

A generalization of `gen` is

`egen`

This command allows to use such operation as `sum, mean, sd,` etc.

Try `help egen`

If you want a simple (sclar, not vector) calculation use `display.` For example, you can compute and display

`display sqrt(16)/ln(16)`

2.1. Random number generation

To start a new STATA session with random number generation issue

`set obs` numberofobs. For example, if you want to have 1000 observation in each variable use

`set obs 1000`

To generate a uniformly distributed RV use

`gen r=uniform()`

To generate a normally distributed $N(0, 1)$ RV use

`gen rn=invnorm(uniform())`

2.2. Data analysis commands

The simplest data analysis command to calculate mean and SD is

`summarize` varname.

To find correlation coefficient use

`corr` var1 var2

To find regression (conditional mean) use

`regress` var1 var2

2.3. Graphs

To plot a histogram use `graph` var, hist

You can specify # of bins and draw a normal density using

`graph` var, `hist bin(4) norm`

To plot a scatterplot use: `graph` var1 var2

To plot boxplot use `graph` var1 var2, `box`

To plot Q-Q plot use `qnorm` var

3. Method of moments, Rice 8.4

There are a few general methods of parameter estimation. Among them is method of moments. The idea of method of moments is to find an estimator by equating sample moment(s) to theoretical moment(s). More precisely, the procedure is as follows (based on the first sample moment):

1. Let $x_1, x_2, ..., x_n$ be iid data drawn from a distribution $F(x;\theta)$ and density function $f(x;\theta)$ where θ is an unknown parameter to be estimated by means of data, i.e.a function of $x_1, x_2, ..., x_n$. Thus an estimator has the form $\widehat{\theta} = \theta(x_1, x_2, ..., x_n)$.
2. We find the theoretical moment as $\int_{-\infty}^{\infty} xf(x;\theta)dx$ which is a function of θ.
3. We find the sample moment as $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
4. We equate the theoretical moment to the sample moment and solve the equation for θ. It gives the MM estimator for θ.

Example. What is the probability that NYC robber will not be caught if he needs 2 minutes to finish off his business, and police arrival after he breaks in follows exponential distribution with parameter λ?

Data

Break in NYC, Bronx	**Time of police arrival after alarm starts, Min**
Sep 1	3.4
Sep 23	2.4
Oct 6	4.2
Oct 18	10.2
Oct 30	7.8
Nov 7	3.1
Dec 13	5.2
Dec 28	7.2
Jan 3	3.9
Jan 30	2.9
Feb 2	9.5

We remember that the density of exponential distribution is $\lambda e^{-\lambda x}$ with the mean

$$E(X) = \frac{1}{\lambda}.$$

So to estimate λ we write the equation

$$\frac{1}{\lambda} = \overline{x}$$

which gives

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}.$$

It is called *harmonic* mean.

Hence, the MM estimator of λ for our data is

$$\widehat{\lambda} = \frac{11}{59.8} = .184.$$

The distribution function can be estimated as

$$1 - e^{-\widehat{\lambda}x} = 1 - e^{-.184x}.$$

If X is the time (min) of police arrival after break in, then the probability not to be caught is

$$\Pr(X > 2) = 1 - \Pr(X \leq 2) = e^{-.184 \times 2} = .7$$

4. Method of maximum likelihood, Rice 8.5

is a very popular general statistical method of estimation.

We start with iid data $x_1, x_2, ..., x_n$. Let they are drawn from a population $f(x;\theta)$. What is the 'observed' density for x_i?

$$f(x_i;\theta)$$

But $x_1, x_2, ..., x_n$ are independent so the observed density for our data is

$$f(x_1;\theta)f(x_2;\theta)...f(x_n;\theta)$$

It is called likelihood. A reasonable estimator of the unknown parameter would maximize this likelihood. In practice we take ln and come to maximization of

$$\sum_{i=1}^{n} \ln f(x_i;\theta)$$

This function is called log-likelihood function of θ (data $x_1, x_2, ..., x_n$ are fixed, given). The maximum likelihood estimator $\widehat{\theta}_{ML}$ maximizes the log-likelihood function. We take derivative of the function with respect to θ to find the maximum.

Example. Maximum likelihood estimation of parameters of normal distribution. The data we have are from normal population

$$x_i \sim N(\mu, \sigma^2)$$

Find MLE for μ and σ^2. The observed likelihood for x_i is

$$f(x_i;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

with the log

$$-\frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(x_i - \mu)^2 - \ln\sqrt{2\pi}$$

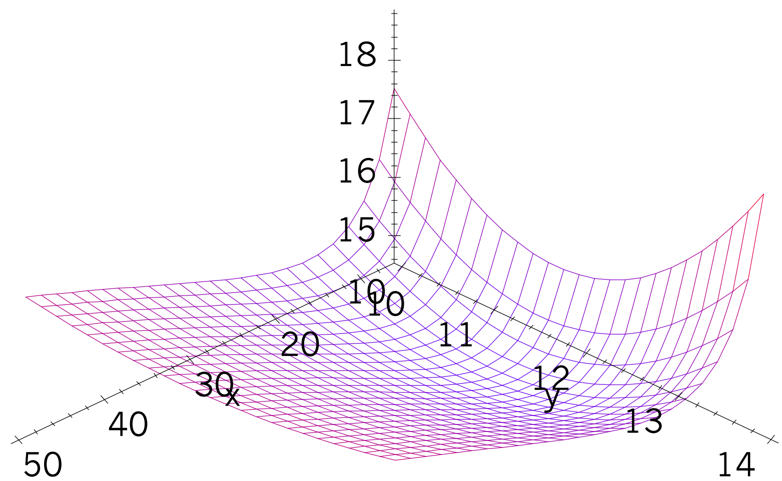
To find the MLE we have to maximize sum of logs (we drop constant term because it does not depend on parameters)

$$
\begin{aligned}
& \sum_{i=1}^{n}\left(-\frac{1}{2}\ln\sigma^2-\frac{1}{2\sigma^2}(x_i-\mu)^2\right) \\
= & -\frac{1}{2}\sum_{i=1}^{n}\left(\ln\sigma^2+\frac{1}{\sigma^2}(x_i-\mu)^2\right) \\
= & -\frac{1}{2}\left(n\ln\sigma^2+\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)
\end{aligned}
$$

Equivalently we can minimize

$$n\ln\sigma^2+\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

for μ and σ^2.



Function of μ and σ^2 to be minimized

Take derivative with respect to μ :

$$\frac{\partial}{\partial\mu}\sum_{i=1}^{n}(x_i-\mu)^2=-2\sum_{i=1}^{n}(x_i-\mu)=0$$

But

$$\sum_{i=1}^{n}(x_i-\mu)=\sum_{i=1}^{n}x_i-n\mu=0$$

which yields

$$\widehat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Arithmetic mean is equal to MLE.

Now we take derivative with respect to σ^2 :

$$\begin{aligned} &\frac{\partial}{\partial \mu}\left(n \ln \sigma^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{ML})^2\right) \\ = \ &\frac{n}{\sigma^2} - \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{ML})^2 = 0 \end{aligned}$$

and we obtain

$$\widehat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{ML})^2.$$

The biased version of the variance is equal to MLE. The unbiased version is equal to MM estimator.

4.1. Optimal asymptotic properties of MLE

MLE has optimal asymptotic properties, i.e. when the number of observations (sample size) is large, $n \to \infty$:

- *Consistency and unbiasdeness*: MLE converges in probability to the true value when $n \to \infty$. This is not the convergence in the sense of calculus but probability theory.

- *Normality*: the distribution of MLE is approximately normal,

  $$\frac{\widehat{\theta}_n - \theta}{\sqrt{var(\widehat{\theta}_n)}} \simeq N(0, 1).$$

  where $var(\widehat{\theta}_n)$ is called the asymptotic variance of MLE.

- *Asymptotic variance* of MLE is calculated as follows

  $$var(\widehat{\theta}_n) = \frac{1}{nI(\theta)}$$

  where $I = I(\theta)$ is called *Fisher information* about θ, and is defined as the negative expectation of the second derivative of the log density:

  $$I(\theta) = -E\left[\frac{d^2}{d\theta^2}\ln f(x;\theta)\right]$$

- *Efficiency:* MLE has minimum variance among all estimators of the true parameter in large samples.

Problem. Find MLE for λ in exponential distribution, asymptotic variance and asymptotic 95% CI.

Solution. The density of exponential distribution is

$$f(x;\lambda) = \lambda e^{-\lambda x}.$$

If x_i are iid with the density $f(x;\lambda)$ then the log-likelihood function is

$$\begin{aligned} \sum_{i=1}^{n} \ln f(x_i;\lambda) &= \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda x_i}\right) = \sum_{i=1}^{n} (\ln\lambda - \lambda x_i) \\ &= n\ln\lambda - \lambda\sum_{i=1}^{n} x_i \end{aligned}$$

To find the MLE for λ we need to maximize the log-likelihood function

$$\sum_{i=1}^{n} \ln f(x_i;\lambda) \Rightarrow \max_{\lambda}$$

that translates into

$$n\ln\lambda - \lambda\sum_{i=1}^{n} x_i \Rightarrow \max_{\lambda}$$

We take derivative and put it zero:

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

which yields

$$\widehat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}.$$

What is the variance of $\widehat{\lambda}_{ML}$? Calculate the Fisher information

$$I(\theta) = -E\left[\frac{d^2}{d\theta^2}\ln f(x;\theta)\right].$$

We have

$$\ln f(x;\lambda) = \ln\lambda - \lambda x$$

and

$$\begin{aligned} \frac{d}{d\lambda}\ln f(x;\lambda) &= \frac{1}{\lambda} - x, \\ \frac{d^2}{d\lambda^2}\ln f(x;\lambda) &= -\frac{1}{\lambda^2} \end{aligned}$$

The information about λ is

$$I(\lambda) = -E\left(-\frac{1}{\lambda^2}\right) = \frac{1}{\lambda^2}.$$

The asymptotic variance is

$$var(\widehat{\lambda}_{ML}) = \frac{1}{nI(\lambda)} = \frac{\lambda^2}{n}$$

For large n

$$\widehat{\lambda}_{ML} \simeq N\left(\lambda, \frac{\lambda^2}{n}\right)$$

Therefore, 95% CI is

$$\begin{aligned} & \left(\widehat{\lambda}_{ML} - 1.96\frac{\widehat{\lambda}_{ML}}{\sqrt{n}}, \widehat{\lambda}_{ML} + 1.96\frac{\widehat{\lambda}_{ML}}{\sqrt{n}}\right) \\ = & \left(\frac{\sqrt{n} - 1.96}{\sqrt{n}}\widehat{\lambda}_{ML}, \frac{\sqrt{n} + 1.96}{\sqrt{n}}\widehat{\lambda}_{ML}\right). \end{aligned}$$

It may be proved that the regression coefficients (intercept and slope) are MLE if the distribution is normal.

5. Linear model. Method of least squares, Rice 14

is applied to estimate parameters of statistical relationship. For example, let y be the weight and x be the height of the person. We expect that y and x are positively correlated, particularly,

$$y = \alpha + \beta x + \varepsilon$$

where ε is residual. If i is a person with measurements (x_i, y_i) then

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

x is called regressor, predictor, or explanatory (independent) variable. y is called 'dependent variable'. ε is called 'residual or error term'.

We assume that

$$E(\varepsilon_i) = 0, var(\varepsilon_i) = \sigma^2.$$

This model is called *linear model* – one of the most popular statistical model to study stochastic relationship.

What is the best estimator for parameters α, β? We find estimators for α, β as those which minimize sum of squares (method of *least squares*)

$$S(\sigma, \beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 .$$

Geometrical interpretation..

Minimum occurs where the derivative is zero:

$$\begin{aligned} \frac{\partial S}{\partial \alpha} &= -2\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0 \\ \frac{\partial S}{\partial \beta} &= -2\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) x_i = 0 \end{aligned}$$

This pair of equations determine estimator for α and β. They are called normal equations. The solution is

$$b = \frac{\sum_{i=1}^n (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

– the estimator of the slope,

$$a = \overline{y} - b\overline{x}$$

– the estimator of the intercept. Regression goes through the mean point.

They are called Ordinary Least Squares (OLS) estimators.

Residuals are $e_i = y_i - a - bx_i$.

OLS leads to the same formula as derived before based on sampling moments for conditional mean $b = \widehat{\sigma}_{xy}/\widehat{\sigma}_x^2$.

Three methods coincide: sampling moments for conditional mean, MLE, OLS.

It is possible to prove that estimators a and b are Best Linear Unbiased Estimators (BLUE). In particular,

$$E(b) = \beta, \quad E(a) = \alpha.$$

5.1. Residual sum of squares and variances of parameters

In conditional mean interpretation $\sigma_{Y|X}^2 = \sigma^2$ is conditional variance. It is an unknown parameter. It can be estimated via Residual Sum of Squares

$$\begin{aligned} SS &= \sum_{i=1}^n (y_i - a - bx_i)^2 \\ &= \sum_{i=1}^n e_i^2 \end{aligned}$$

and the estimator is

$$\widehat{\sigma}^2 = \frac{SS}{n-2}$$

We have $n-2$ in denominator to adjust for degrees of freedom (2 degrees of freedom spent for estimation α and β). It is possible to show that $\widehat{\sigma}^2$ is an unbiased estimator of σ^2.

Parameters a and b are also random variables and they the variance. It gives you the measure how well parameters are estimated:

$$var(b) = \frac{\widehat{\sigma}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

and

$$SD(b) = \sqrt{var(b)} = \frac{\widehat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}.$$

Less $SD(b)$ better estimation is. $SD(b)$ tells us how accurate could we estimate the slope.

5.2. Coefficient of determination, R^2

Is regression adequate? How to measure the quality of relationship? Correlation coefficient (c.c) is a measure of statistical relationship between x and y. Coefficient of determination is squared c.c

$$R^2 = r^2.$$

It has a nice interpretation due to the following equality

$$\begin{aligned} R^2 &= 1 - \frac{\sigma^2_{Y|X}}{\sigma^2_Y} = \frac{\sigma^2_Y - \sigma^2_{Y|X}}{\sigma^2_Y} \\ &= 1 - \frac{SS}{\sum(y_i - \overline{y})^2} \end{aligned}$$

We can interpret $\sigma^2_{Y|X}$ as unexplained variance (residual SS) and σ^2_Y is total variance $\sum(y_i - \overline{y})^2$ is total sum of squares.

Adjusted C.D. is when unbiased versions of variances are used, i.e.

$$R^2_{adj} = 1 - \frac{SS/(n-m)}{\sum(y_i - \overline{y})^2/(n-1)}.$$

Therefore, R^2 is the portion of variation in y we can explain by x. For example, if c.c =.8 then $R^2 = .8^2 = .64$ means that 64% of $y-$variance may be explained by regressor.

Properties of R^2 :

1. R^2 is nonnegative and less than 1.

2. If $R^2 = 1$ then SS=0 and all points (x_i, y_i) lies on a straight line.

3. R^2 does not change when x and y are measured on another scale (invariant under linear transformation).4.

4. If R^2 close to 0 then b is close to zero. Regression is constant, no slope.

Problem. Generate two RVs with the regression slope 1/2.

Solution. We use the formula

$$\beta = \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}^2_x}$$

We can take $\widehat{\sigma}^2_x = 1$. Thus we need to generate two RVs with covariance 1/2. Let us generate $x \sim N(0,1)$ and $y \sim N(0,1)$ independent of x. Let t be any constant, I take $z = y + tx$ and consider the covariance

$$cov(x,z) = cov(x, y+tx) = cov(x,y) + t \times cov(x,x) = t \times var(x) = t.$$

Therefore, I take $t = 1/2$.

STATA program:

5.3. Predicted values and possible extensions

Once regression is estimated one can predict $y-$values given x-values (remember, regression is conditional mean?). The predicted value at $x = x_i$ is

$$\widehat{y}_i = a + bx_i.$$

The residual is for the ith observation is

$$e_i = y_i - \widehat{y}_i.$$

Residuals give you information how far y_i is from the regression line.

Before we dealt with bivariate regression. It is possible to generalize to multivariate regression, e.g.,

$$y_i = a + bx_i + cz_i + \varepsilon_i.$$

Coefficients a, b, c can be estimated again using Least Squares method:

$$\sum_{i=1}^{n}(y_i - a - bx_i - cz_i)^2 \Rightarrow \min_{a,b,c}$$

Use regress in STATA, e.g. regress y x z

Also we can use regression for nonlinear relationship, e.g. quadratic

$$y = a + bx + cx^2 + \varepsilon.$$

Then, you set $z = x^2$ and come to multivariate regression

$$y = a + bx + cz.$$

LOG transformation: sometimes relationship is in LOG (multiplicative relationship) as

$$\ln y = a + b \ln x$$

which corresponds to

$$y = Ax^b.$$

Other nonlinear relationships can be analyzed via linear regression:

- y=a/(b+x)
- y=x/(a+bx)
- y=1/(1+ax^b)

6. Homework (due March 3)

1. (5 points). Using STATA verify that the probability .7 in our Robber and Police example is correct. To do this

take $n = 10000$ (`set obs` 10000) and generate 10000 observations from the exponential distribution with $\lambda = .184$ based on the uniform (0,1) distribution. Then calculate # of cases for which $x > 2$. Use command `summarize` polar>2 where polar is the generated variable (10000 observations).

2. (7 points). Show that correlation coefficient is zero if and only if regression line is parallel to x-axis. Verify this using STATA: `set obs 1000`; create two independent normally distributed RVs x and y and `regress y x`. Is the regression line parallel to x-axis? Include STATA output.

3. (5 points). Can two RVs be functionally related with correlation coefficient 0? Give an example/counterexampe using STATA.

4. (3 points). Is the MM estimator of λ for exponential distribution unbiased? Justify the answer.

5. (4 points.) Show that if x_i are iid distributed as $N(\mu, \sigma^2)$ then $\widehat{x}^2$ is a biased estimator of μ^2. Prove mathematically or via simulation using STATA (include output).

6. (5 points). What is probability that the true value of λ in 'Break in' example is more than .2?

7. (7 points). Consider and iid sample from a double exponential distribution with the density $f(x; \lambda) = (2\lambda)^{-1} \exp(-\lambda^{-1}|x|)$. Find the MLE for parameter λ.

8. (6 points). If x_i are iid with known population mean μ but unknown population variance σ^2. What is an unbiased estimator of variance? Justify, please.

9. (6 points). Prove that the coefficient of determination defined as $1 - \sigma^2_{Y|X}/\sigma^2_Y$ is equal to squared correlation coefficient defined as $\sigma_{XY}/(\sigma_X \sigma_Y)$. You can prove mathematically or using a simulated data. Include STATA output.

10. (8 points). Prove that regression line goes through the center of data, $(\overline{x}, \overline{y})$. Verify this statement using STATA: `set obs 100` and create two normally distributed RVs, x and y; `regress y x`; compute y when x=$\overline{x}$. Does it give $\overline{y}$? Include STATA output.

11. (6 points). Prove that the method of least squares applied to estimation of the mean in iid sample leads to the arithmetic mean.