

1. Suppose the observation  $Z$  and the true value  $S > 0$  are linked through *multiplicative* measurement error:

$$Z = S \cdot (\sigma\varepsilon),$$

where random variables  $S$  and  $\varepsilon$  are independent and  $\varepsilon$  has probability density function (pdf)  $f^{(\varepsilon)}$ . Derive the conditional pdf  $f^{(Z|S)}(z|s)$ .

2. Consider a random variable  $X$  with mean  $\mu_X$  and variance  $\sigma_X^2$ , and let

$$Y = g(X),$$

where  $g(\cdot)$  is a twice differentiable function. Using a Taylor series expansion, it can be shown that  $\mu_Y$  and  $\sigma_Y^2$  can be approximated by

$$\mu_Y \simeq g(\mu_X) + g''(\mu_X)\sigma_X^2/2$$

$$\sigma_Y^2 \simeq (g'(\mu_X))^2\sigma_X^2,$$

where  $g'$  and  $g''$  are, respectively, the first and second derivatives of  $g$ . [This approach to finding an approximate relationship between moments of  $X$  and moments of  $Y$  is known as the  $\delta$ -method.]

- (i) Suppose  $X \sim \text{Bin}(n, p)$ , the binomial distribution with parameters  $n$  and  $p$ . Let  $Y = \text{logit}(X/n)$ . Find approximate expressions for  $\mu_Y$ ,  $\sigma_Y^2$ .

Suppose  $\sigma_X$  is in fact a function of  $\mu_X$ . Write:

$$\sigma_X \equiv \sigma(\mu_X).$$

- (ii) Use the  $\delta$ -method to show that the transformation  $g$  in  $Y = g(X)$ , for which

$$\sigma_Y \simeq \text{constant},$$

satisfies  $g'(x) = \text{constant}/\sigma(x)$ . [The resulting  $g$  is known as a variance-stabilizing transformation.]

- (iii) Suppose  $X$  is a positive random variable and

$$\sigma_X = a(\mu_X)^b.$$

Show that

$$g(x) = c_1 x^{1-b} + c_2$$

is a variance-stabilizing transformation.

3. The `data.frame`, `air`, in S-Plus contains 111 observations (rows), and 4 variables (columns), taken from an environmental study that measured the four variables, ozone, solar radiation, temperature, and wind speed, for 111 consecutive days in New York. Ozone is the only variable of interest in (a)-(d).
- (a) Obtain summary information on the data by using the `summary()` function.
  - (b) Make a scatter plot of ozone versus temperature and label the x-axis with ‘Temperature (F)’, the y-axis with ‘Ozone (ppm)’ and put on the title ‘The NY ozone data’. Don’t print this plot.
  - (c) Repeat the same plot command used in (b) but with `type=‘n’` (plot nothing). Then (by using the `points()` function to add to the plotting device):
    - (i) Plot ozone versus temperature using plot symbol number 1 (`pch=1`), but only use the observations where wind is less than its median.
    - (ii) Plot ozone versus temperature using the plot symbol number 2 (`pch=2`), but now use the observations where wind is greater, or equal, than its median.
  - (d) Add a horizontal line to the plot showing ozone’s median (`abline()`) and a vertical line showing temperature’s median. Use line type 2.

Bonus: before you print out your plot, execute these two commands:

```
> legend(57,5.518,c('low wind','high wind'),marks=c(1,2))
> lines(lowess(air$te,air$oz),lty=3) ## smooth trend line
```

Hand in the final plot made and the commands used to create it.