HG
Oct. 06

ECON 4130

Supplementary Exercises 7 - 8

Exercise 7

Let X be the number of traffic accidents occurring during t months in a region. Assume that X is poisson distributed with parameter λt (i.e., $X \sim \text{pois}(\lambda t)$).

a. Explain why the parameter λ can be interpreted as a theoretical incidence rate, i.e., the expected number of accidents per month.

b. We cannot observe λ directly, but we can observe X instead. Show that the estimator $\hat{\lambda} = X/t$

i) is unbiased for all t,

ii) is consistent as $t \to \infty$ (use Chebyshev's inequality).

c. Using the fact that X is approximately normally distributed when λt is large (≥ 10 is usually considered sufficient), develop an approximate $1-\alpha$ confidence interval (CI) for λ based on X.
[**Hint:** Show first, using Slutsky's lemma, that

$$\sqrt{t}\,\frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}}} \underset{t\to\infty}{\overset{D}{\rightarrow}} Z \sim N(0,1)$$

Note also that $\sqrt{t}\,\dfrac{\hat{\lambda}-\lambda}{\sqrt{\lambda}} = \dfrac{X-\lambda t}{\sqrt{\lambda t}}$]

Exercise 8

We are interested in the monthly incidence rate of traffic accidents in Norway. From Statistical Office Norway (SSB), we obtain the number of traffic accidents registered in the period 2003 – 2005, as given in table 1,

Table 1

Year	No. of traffic accidents
2003	8266
2004	8425
2005	8078
Sum	24769

We want a 95% CI for the monthly incidence rate based on these results. Let X_1, X_2, X_3 denote the rv's behind the three observations in table 1. Our first approach is to calculate a "t-interval" for the incidence rate, called λ , based on the following model

Model 1 X_1, X_2, X_3 are *iid* with $X_i \sim N(\mu, \sigma^2)$ where $\mu = 12\lambda$

[**Hint:** When $X_1, X_2, \ldots, X_n$ are *iid* with $X_i \sim N(\mu, \sigma^2)$, we remember from the basic statistic course that an (exact) $1-\alpha$ CI for μ (the so called "t-interval") is

$$\bar{X} \pm t_{1-\alpha/2,\ n-1} \cdot \frac{S}{\sqrt{n}}, \qquad \text{where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{, and}$$

where $t_{1-\alpha/2,\ n-1}$ is the $1-\alpha/2$ percentile in the t-distribution with $n-1$ degrees of freedom.]

a. Calculate the 95% CI for μ based on model 1 and transform the interval to a corresponding CI for λ. Explain why the interval for λ must have the same degree of confidence as the one for μ . Discuss briefly whether the assumptions in model 1 appear reasonable or not.

b. An alternative approach is to assume that the total number of registered accidents in the period 2003 - 2005, $X = X_1 + X_2 + X_3$, is poisson distributed, i.e.,

Model 2 $X \sim \text{pois}(t\lambda)$ with $t = 36$

Calculate an approximate 95% CI for λ based on model 2, and compare with the CI in **a**.

c. One reasonable criticism that can be raised against the model 2, is the apparently unrealistic assumption of *constant incidence rate* that underlies the poisson model, which, among other things, implies that all the months of the year have the same incidence rate, λ. For example, table 2, that gives the number of accidents for January and June, appears to support this criticism.

Table 2

	Number of accidents	
Year	January	June
2003	576	805
2004	616	847
2005	588	853

Luckily, the poisson model offers an easy way to accommodate this criticism. To see this, first prove the following result [**Hint:** Use the mgf for the poisson distribution]:

Property 1 Let $Y_1, Y_2, \ldots, Y_k$ be independent and poisson distributed with $Y_j \sim \text{pois}(\mu_j)$ for $j = 1, 2, \ldots, k$. Then $Y = Y_1 + Y_2 + \cdots + Y_k \sim \text{pois}(\mu)$ where $\mu = \mu_1 + \mu_2 + \cdots + \mu_k$.

d. In order to accommodate the criticism, we suggest the following model: Let Y_{ij} be the number of accidents in month *j* ($j = 1, 2, \ldots, 12$) in year *i* (i = 1,2 3). Assume

Model 3 The Y_{ij}'s are independent and poisson distributed with $Y_{ij} \sim \text{pois}(\lambda_j)$ for $j = 1, 2, \ldots, 12$ and $i = 1, 2, 3$.

Show that $X = \sum_{i=1}^{3}\sum_{j=1}^{12} Y_{ij} \sim \text{pois}(36\bar{\lambda})$ where $\bar{\lambda} = \frac{1}{12}\sum_{j=1}^{12}\lambda_j$ is the average monthly incidence rate.

e. Show that the estimator in exercise 7, $\hat{\lambda} = \frac{X}{t}$, where $t = 12r$ is the number of months and *r* the corresponding number of years, is unbiased for $\bar{\lambda}$ and with variance, $\text{var}(\hat{\lambda}) = \frac{\bar{\lambda}}{t}$. This shows that $\hat{\lambda}$ also is consistent for $\bar{\lambda}$ as $t \to \infty$ (why?).

f. Explain why the CI in **b** is still valid, but now as an approximate 95% CI for the new parameter, $\bar{\lambda}$.