Mathematical statistics and probability

Einar Wilhelm Bratthall

January 2022

1 Innledning til sannolikhetsteorin

2 Sannolikhetstorins grunder

3 Endimensionella stokastiska variabler

4 Flerdimensinoaella stokastiska variabler

5 Expected value

6 Normal distribution

7 Binomial distribution

8 Random numbers and simulation

9 Introduction to statistical theory

10 Descriptive statistics

11 Point approximation

This is a common thing in daily life, for instance, you approximate someones age by how they look.

11.1 Political survey example

You can approximate the opionons of a population by asking a small sample. Lets say you ask 1000 people and $x = 350$ say yes to a yes and no question. $P^*_{obs} = 350/1000$. This would be $Hyp(N, 1000, p)$ but as N is very large (5 million) it can be approximated with $Bin(1000, p)$.

$p^* = X/1000$ where X is the count of YES and X is Bin(1000,p).

$$E(p^*) = E(\frac{X}{1000}) = \frac{E(X)}{1000} = p$$

$$V(p*) = V(\frac{X}{1000}) = \frac{V(X)}{1000^2} = \frac{p(1-p)}{1000}$$

$$D(p^*) = \sqrt{V(X)}$$

Inserting the values gives a standard deviation of 1.5 percent

11.2 General formula

The goal with a point approximation is to apprixmate a paramter θ. We find θ by obesrvations x_i are events from the s.v X_i. θ^* will change with each set of θ^*_{obs}, in fact each θ^*_{obs} is an outcome of θ^*. The definition is a bit circular, as each X_i depens on θ^*.

Definition: a point approximation θ^*_{obs} is said to be ”väntevärdesriktig” if

$$E(\theta^*) = \theta \ \forall \ \theta \in \Omega_\Theta$$

where Ω_Θ is ”utfallsrummet”.

When the sample size $n \to \infty$: $\theta^*_{obs} \to \theta$

Definition: ”Medelvärdekvadratfelet” MSE is

$$MSE = E((\theta^* - \theta)^2)$$

where θ^* is the ”stickprovsvariabel” and the systematic error is $E(\theta^*) - \theta$

11.3 Approximating expected value variance

One can approximate μ with $\mu^*_{obs} = \overline{x}$.
Approximating the variance can be done by the formula

$$(\sigma^2)^*_{obs} = s^2 = \frac{1}{n-1}\sigma^n_{i=1}(x_i - \overline{x})^2.$$

Both of these are ”väntevärdesriktig”

11.4 Maximum-likelihood-method

X is poisson fördelad

$$X \in Po(\mu)$$

$$p_x(k) = \frac{\mu^k}{k!} e^{-\mu}$$

It seems to be to simply calcualte the probability of your data, and maximize that probabiliy based on a standard model. If we have 2 poisson phone calls $X_1, X_2 = 10, 12$ and they are independent $Po(\theta)$ then their probabily is

$$P(X_1 = 10, X_2 = 12) = \frac{\theta^{10}}{10!} e^{-\theta} * \frac{\theta^{12}}{12!} e^{-\theta} = \frac{\theta^{10+12}}{10!12!} e^{-2\theta}$$

the θ that gives the maximal p is the solution. In this case it can easily be shown that it is $\theta^*_{obs} = \overline{x}$ A good technique can be to use "logaritmering" and the take the derivative and solve for 0.

11.5 Smallest-Square-Method

12 Interval approximation

13 Testing hypothesis

14 Regression analysis