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

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

$$V(\mu^*) = \frac{\sigma^2}{n}$$

Both of these are "väntevärdesriktig"
Example:

Given s.v. X with some $E(X) = \mu = \theta$ and some σ that's unknown. We know the number of trials N, $\overline{x}$ and s. We can rewrite this as

$$D(X) = \sqrt{V(X)} = \sqrt{V(\mu*)} = \sqrt{\frac{\sigma^2}{N}} = \sqrt{\frac{s^2}{N}}$$

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

We have X_I describing x_i with known expected value functions only dependent on one paramater. $E(X_i) = \mu_i(\theta) + \epsilon_i$ where ϵ descirbes the error from given experimetn. Then minimizing the function

$$Q(\theta) = \Sigma_{i=1}^n (x_i - \mu_i(\theta))^2 \tag{1}$$

gives the MK-approximation for the s.v. This value is called θ^*_{obs} If all functions $\mu_i(\theta)$ are equal Q will be minimized when $\mu(\theta) = \overline{x}$ giving. $\theta^*_{obs} = \mu^{-1}(\overline{x})$ The more general method is solving

$$\frac{dQ(\theta)}{d\theta} = \sum .... \tag{2}$$

11.6 Tillämpning på normalfördelningen

12 Interval approximation

Insead of directly approximation the expected value, one can give an interval with some probability that the expected value lies within this interval. It is best shown in an example:
Lets say you measure something with unknown $E(X) = \theta$ and get the approximation x. The measurement error is $N(0, \sigma)$ and σ is known. Then x is an

obesrvation of X where $X \in N(\theta, \sigma)$, here an interval can be found with some probability. For instance with 95 percent confidence.

$$\theta - 1.96\sigma < X < \theta + 1.96\sigma$$

or

$$X - 1.96\sigma < \theta < X + 1.96\sigma$$

The interval becomes $I_\theta = (x - 1.96\sigma, x + 1.96\sigma)$

In some examples prorperties from chapter 6 are very useful here.

If X_i is independent $N(\mu, \sigma)$ and $\overline{X}$ is known then

$$\overline{X} \in N(\mu, \frac{\sigma}{\sqrt{N}})$$

12.1 Tillämpad på normalfördelningen

13 Testing hypothesis

We have a hypotheiss and we are looking to test it.

14 Regression analysis