Západočeská univerzita v Plzni

Matematické modely v Ekonometrii

KMA/MME

1. semestrální práce

Autor:
Marek Lovčí

October 26, 2020

FAKULTA
APLIKOVANÝCH VĚD
ZÁPADOČESKÉ
UNIVERZITY
V PLZNI

1 Zadání

Pomocí numerických experimentů ověřte základní statistické vlastnosti odhadů klasického a klasického normálního modelu uvedených v Gauss-Markovově větě. Zaměřte se například na ověření:

- nestrannosti odhadů parametrů $\beta_0, \beta_1, \ldots$;
- nestrannosti odhadu σ^2;
- normality odhadů parametrů pro klasický normální regresní model.

2 Teorie

Polynomiální regrese prvního stupně, známa též jako lineární regrese je jednou ze základních matematických metod. Výstupem metody je proložení souboru vstupních dat (bodů) v n-dimenzionálnním prostoru přímkou. Díky své jednoduchosti a rychlosti výpočtu nachází tato metoda uplatnění v mnoha oborech, např. strojovém učení.

Přímkový regresní model je vyjádřen následujícím vztahem.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i;\ i = 1, 2, \ldots, n \tag{1}$$

Odhad lineárního modelu je možné provést několika různýmy způsoby, nejčastěji však metodou nejmenších čtverců (*Mean Squared Error*).

$$\widehat{MSE}(\beta_0, \beta_1) \equiv \frac{1}{n} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 \tag{2}$$

Při odhadu parametrů provádíme minimalizaci ztrátové funkce.

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2 \tag{3}$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2 \tag{4}$$

Dle teorie přednášek (konkrétně přednáška 2, část 2) předpokládáme, že odhad parametrů metodou nejmenších čtverců je nejlepší nestranný lineární odhad (*Best Linear Unbiased Estimator*). Nestrannost odhadů parametrů znamená, že očekávaná hodnota (tj. střední hodnota - *expected*) odhadů je rovna hodnotě parametrů samotných.

$$E(b_i) = \beta_i \tag{5}$$

Zajímavý pohled na MNČ jako na funkci lineární kombinace parametrů θ.

$$\theta = a_0 \beta_0 + a_1 \beta_1 \tag{6}$$

$$\hat{\theta} = a_0 b_0 + a_1 b_1 \tag{7}$$

$\hat{\theta}$ je odhadem regresní přímky, $a_0 = 1$, $a_1 = x$ jsou libovolná reálná čísla. Potom lze uvažovat MSE jako rozptyl veličiny $\hat{\theta} - \theta$.

$$MSE(\hat{\theta}) \triangleq \mathbb{E}\left[(\hat{\theta} - \theta)^2\right] \tag{8}$$

Střední kvadratická chyba je tudííž vždy nezáporná a nulová pouze v případě, kdy je odhad bezchybný. Vlastností takto definované střední kvadratické chyby je, že se dá rozložit na součet systematické chyby (vychýlení) a rozptylu odhadu (náhodná nepřesnost odhadu).

$$MSE(\hat{\theta}) = Bias(\hat{\theta})^2 + Var(\hat{\theta}) \tag{9}$$

Jestliže budeme požadovat konstrukci intervalových odhadů a testů hypotéz, tak bude třeba přidat předpoklad normality dat pro náhodný vektor. V tom případě má náhodná veličina $\hat{\theta}$ normální rozdělení pravděpodobnosti s následující střední hodnotou a rozptylem.

$$E(\hat{\theta}) = a_0\beta_0 + a_1\beta_1 \tag{10}$$

$$var(\hat{\theta}) = \sigma^2 v^2(a_1, a_2) \tag{11}$$

3 Vypracování

V rámci vypracování počítáme dolní a horní odhady pro intervaly spolehlivosti. Výsledky jsou zaznamenány na grafu 1. Pouhýh několik výsledků z evšech provedených simulací se nachází mimo očekávané hodnoty.

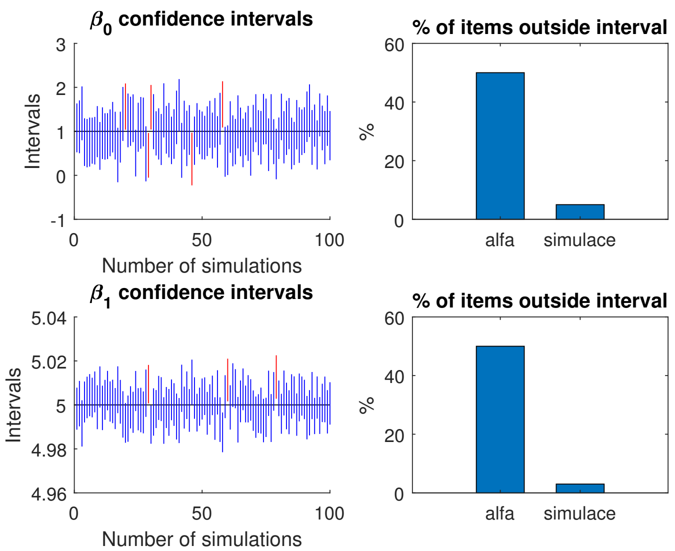

Figure 1: Vývoj dolních a horních odhadů parametrů β_i

Na následujícím grafu lze zhodnotit, že četnosti odhadů parametrů β_i mají Gaussovo rozložení.

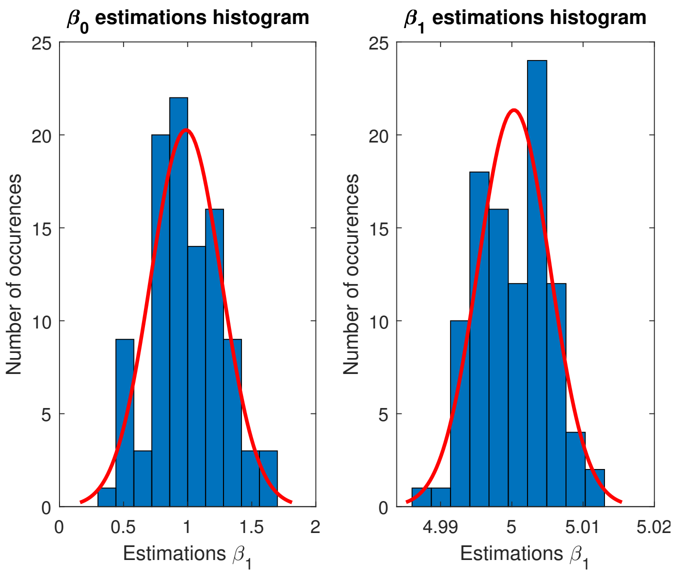

Figure 2: Četnosti hodnot β_0, β_1 napříč simulacemi.

Na posledním grafu je znázorněna konvergence odhadů k teoretickým hodnotám pro zvyšující se počet dat.

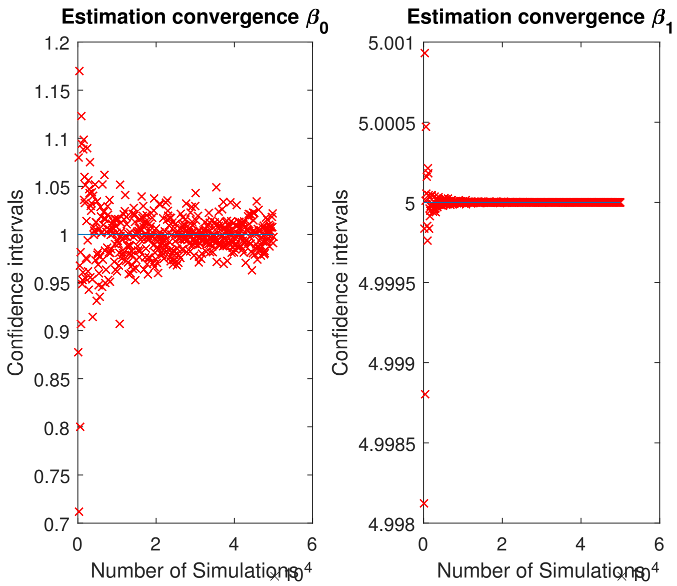

Figure 3: Konvergence odhadů k teoretickým hodnotám v závislosti na rostoucím počtu dat.

4 Závěr

V semestrální práci jsme ověřili několik statistických vlastností lineární regrese. Poznatky zjištěné experimentálně odpovídají teorii.