General notes

Midterm will contain 6 problems. All the problems have equal weights. Midterm duration is 120 minutes. Closed book, one A4 cheatsheet is allowed.

Demo variant «Dolly»

1. Let X be a random variable that follows a uniform distribution on the interval $[10, 20]$ and $g(X) = \frac{1}{X}$.

   (a) Find $\mathbb{E}(g(X))$ and $\mathbb{V}\text{ar}(g(X))$ exactly by computing the expectation and variance directly from the definition of expectation.

   (b) Use the delta method to approximate $\mathbb{E}(g(X))$ and $\mathbb{V}\text{ar}(g(X))$.

   (c) Compare the exact variance obtained in part (a) with the approximation from part (b). Discuss the accuracy of the delta method in this case.

2. Consider a logistic regression model for the probability of success in the Econometrics course:

$$\mathbb{P}(y = 1 \mid h) = \frac{\exp(\beta_0 + \beta_1 h)}{1 + \exp(\beta_0 + \beta_1 h)},$$

   where y is a binary outcome variable, and h is a number of drawn hedgehogs.

   Suppose that the model has been estimated using a dataset of $n = 1000$ students, and the estimated coefficients along with their standard errors are: $\hat{\beta}_0 = -2.5$ with $se(\hat{\beta}_0) = 0.5$ and $\hat{\beta}_1 = 1.2$ with $se(\hat{\beta}_1) = 0.3$.

   (a) Compute the predicted probability $\hat{p}$ for $h = 2$.

   (b) Use the delta method to approximate the variance of $\hat{p}$ for $h = 2$.

   (c) Construct an approximate 95% confidence interval for p using a normal approximation and delta method.

   (d) Discuss the limitations of using the delta method in this context.

3. The dataset contains 1000 observations. We have estimated logistic regression A:

$$\hat{\mathbb{P}}(y_i = 1 \mid a_i, b_i, c_i, d_i) = \Lambda(0.3 + 0.1a_i + 0.2b_i - 0.3c_i + 0.2d_i), \quad \ln L = -330,$$

   and logistic regression B:

$$\hat{\mathbb{P}}(y_i = 1 \mid a_i, b_i, c_i, d_i) = \Lambda(0.1 + 0.4a_i), \quad \ln L = -335.$$

   Here $\ln L$ denotes the maximal value of the log-likelihood function.

   (a) Compare these two nested models using LR-test. Use 5% significance level. Clearly state H_0, H_a, the distribution of the test statistic under H_0 and critical region.

   (b) Compare these models using corrected Akaike information criterion.

   (c) Calculate $\ln L$ for the trivial model $\hat{\mathbb{P}}(y_i = 1 \mid a_i, b_i, c_i, d_i) = \Lambda(\hat{\beta}_0)$ given that $y_i = 1$ in 30% of the observations.

4. The dataset contains 1000 observations. Winnie-the-Pooh has estimated logistic model for the probability that honey is good ($h_i = 1$). Predictor x_i is the height of the tree and $b_i = 1$ for good bees and $b_i = 0$ for bad bees.
$$\hat{\mathbb{P}}(h_i = 1 \mid x_i, b_i) = \Lambda(0.2 + 0.1x_i + 0.2b_i).$$
   (a) For which height of the tree the marginal effect $\partial\hat{P}/\partial x$ is maximal given that bees are good?
   (b) Draw the region of the plane where the predicted probability is from 0.2 to 0.4.
   (c) For which height of the tree the partial effect of bees type (good bees minus bad bees) is maximal?

5. Consider the simultaneous equation model with endogeneous variables v_i and w_i.
$$\begin{cases} v_i = \alpha_1 + \alpha_2 x_i + \alpha_3 w_i + \alpha_4 a_i + \alpha_5 b_i + u_{1i} \\ w_i = \beta_1 + \beta_2 v_i + \beta_3 b_i + u_{2i}. \end{cases}$$
   (a) Check the order identification condition for each equation.
   (b) Check the rank identification condition for each equation.

6. Something on logit model from LSE external exam.