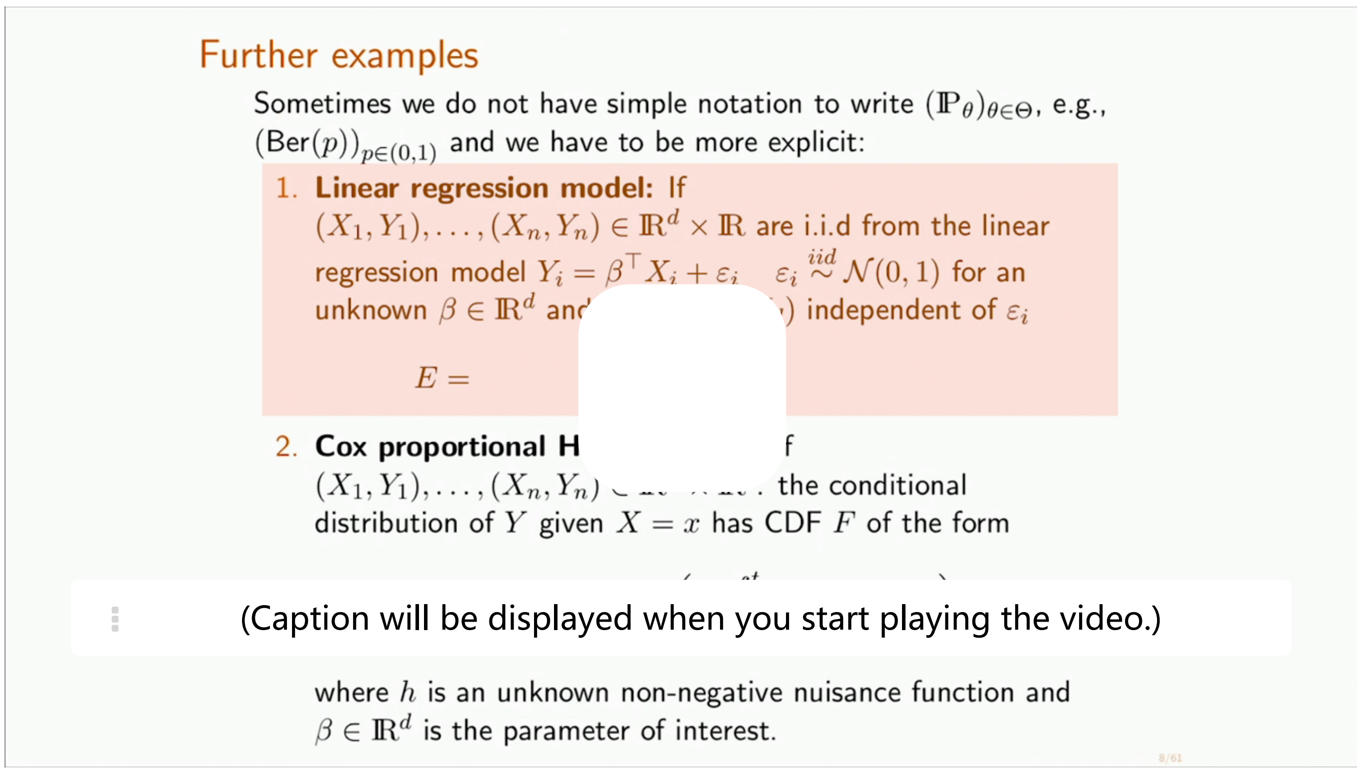
9. Further examples

Linear regression model and Cox proportional Hazard model



Start of transcript. Skip to the end.

There's a couple more examples that I wanted you to look at.

One is linear regression.

So this is something we will come back to.

So you can see also that it's not

entirely trivial to write what the distribution is.

I'll write it on the board for you.

视频

下载视频文件

字幕

下载 SubRip (.srt) file

下载 Text (.txt) file

Linear regression as a statistical model I

2/2 points (graded)

Consider the linear regression model introduced in the slides and lecture, restated below:

Linear regression model : $(X_1,Y_1),\dots,(X_n,Y_n)\in\mathbb{R}^d\times\mathbb{R}$ are i.i.d from the linear regression model $Y_i=\beta^\top X_i+\varepsilon_i,\quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0,1)$ for an unknown $\beta\in\mathbb{R}^d$ and $X_i\sim\mathcal{N}_d(0,I_d)$ independent of ε_i.

Suppose that $\beta=\mathbf{1}\in\mathbb{R}^d$, which denotes the d-dimensional vector with all entries equal to 1.

What is the mean of Y_1?

$\mathbb{E}[Y_1]=$ 0 ✔ Answer: 0

What is the variance of Y_1? (Express your answer in terms of d.)

$\mathsf{Var}(Y_1)=$ d+1 ✔ Answer: d+1

$d+1$

STANDARD NOTATION

Solution:

By definition of the model and setting $\beta=\mathbf{1}$, we have

$$Y_1 = \beta^T X_1 + \varepsilon_1 = \mathbf{1}^T X_1 + \varepsilon_1 = \varepsilon_1 + \sum_{j=1}^{d} X_{1,j}.$$

where $X_{i,j}$ denotes the j'th coordinate of $X_i \sim \mathcal{N}(0, I_d)$. By linearity of expectation,

$$\mathbb{E}[Y_1] = \mathbb{E}[\varepsilon_1] + \sum_{j=1}^{d} \mathbb{E}[X_{1,j}] = 0$$

Next we compute the variance. Since $X_{1,1}, \ldots, X_{1,d}, \varepsilon_i$ are mutually independent, the variance is additive:

$$\mathbf{Var}[Y_1] = \mathbf{Var}[\varepsilon_1] + \sum_{j=1}^{d} \mathbf{Var}[X_{1,j}] = d + 1$$

because $X_{1,1}, \ldots, X_{1,d}, \varepsilon_1 \overset{iid}{\sim} N(0,1)$.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Linear regression as a statistical model II

2/2 points (graded)

Recall the linear regression model as introduced above in the previous question. This model is parametric, although it is not written in the standard notation previously introduced for parametric statistical models. In this problem, you will explicitly write the linear regression model as a parametric statistical model.

We will represent the linear regression model as an ordered pair $(E, \{P_\beta\}_{\beta \in \Theta})$. Here E denotes the sample space associated to the distribution P_β, where P_β is defined as follows for $\beta \in \mathbb{R}^d$:

The random ordered pair $(X, Y) \subset \mathbb{R}^d \times \mathbb{R}$ *is distributed as* P_β *if:*

- $X \sim N(0, I_d)$,
- $Y \sim \beta^T X + \varepsilon$, *where* $\varepsilon \sim N(0,1)$ *and* ε *is independent of* X.

The set Θ in the ordered pair $(E, \{P_\beta\}_{\beta \in \Theta})$ denotes the parameter space for this model.

The sample space for the linear regression model can be written $E = \mathbb{R}^k$ for some integer k. What is k? (Express your answer in terms of d.)

Hint: You should use the fact that $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$ for all integers $m, n \geq 0$.

$k =$ d+1 ✔ **Answer:** d+1

$d + 1$

The parameter space for the model can be written as $\Theta = \mathbb{R}^j$ for some integer j. What is j? (Express your answer in terms of d.)

$j =$ d ✔ **Answer:** d

d

STANDARD NOTATION

提交 你已经尝试了1次（总共可以尝试2次）