

1. Assume $\mathbb{E}[y] < \infty$.

(a) Prove

$$\mathbb{E}[\mathbb{E}[y|x]] = \mathbb{E}[y].$$

$$\begin{aligned} \mathbb{E}[\mathbb{E}[y|x]] &= \int_X \mathbb{E}[y|x] f_x(x) dx && \text{(def of expectation)} \\ &= \int_X \left(\int_Y y f_{y|x}(y|x) dy \right) f_x(x) dx && \text{(expand } \mathbb{E}[y|x], \text{ def of expectation)} \\ &= \int_X \int_Y y (f_{y|x}(y|x) f_x(x)) dy dx && \text{(rearrange)} \\ &= \int_X \int_Y y f_{y,x}(y, x) dy dx && \text{(joint pdf = conditional pdf * marginal pdf)} \\ &= \mathbb{E}[y] && \text{(def of expectation) } \blacksquare \end{aligned}$$

(b) Prove

$$\mathbb{E}[\mathbb{E}[y|x_1, x_2]|x_1] = \mathbb{E}[y|x_1].$$

$$\begin{aligned} \mathbb{E}[\mathbb{E}[y|x_1, x_2]|x_1] &= \int_{X_2} \mathbb{E}[y|x_1, x_2] f(x_2|x_1) dx_2 && \text{(def of expectation)} \\ &= \int_{X_2} \left(\int_Y y f(y|x_1, x_2) dy \right) f(x_2|x_1) dx_2 && \text{(expand } \mathbb{E}[y|x], \text{ def of expectation)} \\ &= \int_{X_2} \int_Y y f(y|x_1, x_2) f(x_2|x_1) dy dx_2 && \text{(rearrange)} \\ f(y|x_1, x_2) f(x_2|x_1) &= \frac{f(y, x_1, x_2)}{f(x_1, x_2)} \cdot \frac{f(x_1, x_2)}{f(x_1)} && \text{(cond'l pdf = joint pdf / marg pdf)} \\ &= \frac{f(y, x_1, x_2)}{f(x_1)} = f(y, x_2|x_1) && \text{(cond'l pdf = joint pdf / marg pdf)} \\ \Rightarrow \mathbb{E}[\mathbb{E}[y|x_1, x_2]|x_1] &= \int_{X_2} \int_Y y f(y, x_2|x_1) dy dx_2 && \text{(substitute into above)} \\ &= \mathbb{E}[y|x_1] && \text{(def of expectation) } \blacksquare \end{aligned}$$

2. (2016 Final: Lecture 3 BLP) Consider a dependent variable y for which

$$\mathbb{E}(y|x) = \beta_2 x^2 + \beta_1 x + \beta_0$$

$$y = \beta_2 x^2 + \beta_1 x + \beta_0 + e$$

where $e \sim N(0, \sigma^2(x))$.

(a) Determine the distribution of y given x .

$$y \sim N(\mathbb{E}(y|x), \sigma^2(x)) = N(\beta_2 x^2 + \beta_1 x + \beta_0, \sigma^2(x))$$

(b) For any $h(x)$ such that $\mathbb{E}|h(x)e| < \infty$, prove the following statements:

$$i) \mathbb{E}[e|x] = 0$$

$$\begin{aligned} \mathbb{E}[e|x] &= \mathbb{E}[y - (\beta_2 x^2 + \beta_1 x + \beta_0)|x] && \text{(Substitute for } e) \\ &= \mathbb{E}[y|x] - \mathbb{E}[(\beta_2 x^2 + \beta_1 x + \beta_0)|x] && \text{(linearity of } \mathbb{E}) \\ &= \mathbb{E}[y|x] - (\beta_2 x^2 + \beta_1 x + \beta_0) && \text{(conditioning thm)} \\ &= \mathbb{E}[y|x] - \mathbb{E}[y|x] = 0 && \text{(substitute) } \blacksquare \end{aligned}$$

$$ii) \mathbb{E}[h(x)e] = 0$$

$$\begin{aligned} \mathbb{E}[h(x)e] &= \mathbb{E}[\mathbb{E}[h(x)e|x]] && \text{(law of iter. expect.)} \\ &= \mathbb{E}[h(x)\mathbb{E}[e|x]] && \text{(conditioning thm)} \\ &= \mathbb{E}[h(x) \cdot 0] = 0 && \text{(substitute) } \blacksquare \end{aligned}$$

Clearly state why the condition $\mathbb{E}|h(x)e| < \infty$ is needed.

- For certain distributions, e.g. Cauchy, we cannot calculate a finite expected value, since the tails do not converge quickly enough to zero, so the conditional mean does not exist. This condition limits our analysis to situations where the distribution of $h(x)e$ does not have this problem, so that we can calculate a finite value for $\mathbb{E}[h(x)e]$.

Do these statements imply that the covariate x is uncorrelated with the (conditional expectation function) error e ?

- Yes; since the expected value of e given x is a constant, then the two variables are mean-independent, which implies e is uncorrelated with any function of the regressor(s) x .

(c) We have shown (in class) that $\beta_2 x^2 + \beta_1 x + \beta_0 := m(x)$ is the predictor of y that minimizes the mean-squared prediction error. Consider predicting e^2 and write the mean-squared error of a predictor $g(x)$. Show that $\sigma^2(x)$ minimizes this mean-squared error.

$$\begin{aligned} \mathbb{E}[(e^2 - g(x))^2] &&& \text{(predictor)} \\ \mathbb{E}[(e^2 - g(x))^2] &= \mathbb{E}[(e^2 - \sigma^2(x) + \sigma^2(x) - g(x))^2] && \text{(special zero)} \\ &= \mathbb{E}[((e^2 - \sigma^2(x)) + (\sigma^2(x) - g(x)))^2] && \text{(group terms)} \\ &= \mathbb{E}[(e^2 - \sigma^2(x))^2] + 2\mathbb{E}[(e^2 - \sigma^2(x))(\sigma^2(x) - g(x))] + \\ &\quad \mathbb{E}[(\sigma^2(x) - g(x))^2] && \text{(expand)} \end{aligned}$$

If we let $g(x) = \sigma^2(x)$:

$$\begin{aligned} \mathbb{E}[(e^2 - \sigma^2(x))^2] &= \mathbb{E}[(e^2 - \sigma^2(x))^2] + 2\mathbb{E}[(e^2 - \sigma^2(x))(\sigma^2(x) - \sigma^2(x))] + \\ &\quad \mathbb{E}[(\sigma^2(x) - \sigma^2(x))^2] && \text{(substitute)} \\ &= \mathbb{E}[(e^2 - \sigma^2(x))^2] + 0 + 0 && \text{(cancel terms)} \end{aligned}$$

$$\Rightarrow \mathbb{E}[(e^2 - g(x))^2] \geq \mathbb{E}[(e^2 - \sigma^2(x))^2] \quad \blacksquare$$

Therefore, mean-squared error of predictor for e^2 is minimized when $g(x) = \sigma^2(x)$.

3. Let $g(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$ be a convex function.

(a) For any random vector x , if $\mathbb{E}\|x\| < \infty$ and $\mathbb{E}|g(x)| < \infty$, prove (Jensen's Inequality)

$$g(\mathbb{E}[x]) \leq \mathbb{E}[g(x)]$$

Let $h(x) = a + bx$ represent a line tangent to $g(x)$ at $x = \mathbb{E}[x]$. Since $g(x)$ is convex, then for all x , $h(x) \leq g(x)$, and $h(\mathbb{E}[x]) = a + b\mathbb{E}[x] = g(\mathbb{E}[x])$.

$$\begin{array}{ll}
g(x) \geq h(x) & (g(x) \text{ is concave}) \\
\mathbb{E}[g(x)] \geq \mathbb{E}[h(x)] & (\text{inequality preserved}) \\
\mathbb{E}[g(x)] \geq \mathbb{E}[a + bx] = a + b\mathbb{E}[x] & (\text{substitute, linearity}) \\
\mathbb{E}[g(x)] \geq g(\mathbb{E}[x]) & (\text{substitute}) \blacksquare
\end{array}$$

(b) With $m = 1$, use Jensen's Inequality to bound $(\mathbb{E}[x])^2$.

$$\begin{array}{ll}
(\mathbb{E}[x])^2 \Rightarrow g(\cdot) = (\cdot)^2 & \\
g(\cdot) \text{ is convex} & (\text{by inspection}) \\
\mathbb{E}[g(x)] \geq g(\mathbb{E}[x]) & (\text{Jensen's ineq}) \\
\mathbb{E}[x^2] \geq (\mathbb{E}[x])^2 & (\text{substitute}) \blacksquare
\end{array}$$

By Jensen's inequality, $(\mathbb{E}[x])^2$ must be less than $\mathbb{E}[x^2]$.

(c) For any random vectors (y, x) , if $\mathbb{E}[|y|] < \infty$ and $\mathbb{E}[|g(y)|] < \infty$, prove (Conditional Jensen's Inequality)

$$g(\mathbb{E}[y|x]) \leq \mathbb{E}[g(y)|x]$$

Similar to 3a above, but consider slices of $g(y, x)$ for any given value of x , i.e. $g(y)|x$. Let $h(y)|x = a + b * y|x$ represent a line tangent to $g(y)|x$ at $y = \mathbb{E}[y|x]$. Since $g(y, x)$ is convex, then for all y at a given x , $h(y)|x \leq g(y)|x$, and $h(\mathbb{E}[y|x]) = a + b\mathbb{E}[y|x] = g(\mathbb{E}[y|x])$.

$$\begin{array}{ll}
g(y) \geq h(y) \text{ for any } x & (g(y)|x \text{ is concave}) \\
\mathbb{E}[g(y)|x] \geq \mathbb{E}[h(y)|x] & (\text{inequality preserved}) \\
\mathbb{E}[g(y)|x] \geq \mathbb{E}[a + b * (y|x)] = a + b\mathbb{E}[y|x] & (\text{substitute, linearity}) \\
\mathbb{E}[g(y)|x] \geq g(\mathbb{E}[y|x]) & (\text{substitute}) \blacksquare
\end{array}$$

(d) With $m = 1$, use the Conditional Jensen's Inequality to bound $(\mathbb{E}[y|x])^2$.

Similar to 3b above, by Jensen's inequality, $(\mathbb{E}[y|x])^2$ must be less than $\mathbb{E}[y^2|x]$.

4. (2017 Prelim) You are asked to determine how the conditional mean of a discrete variable y depends on a (continuous) conditioning variable x . With a discrete dependent variable, the assumption about the form of the conditional mean is replaced with an assumption about the entire conditional distribution for y . You need to consider two cases.

Case 1: y takes only 2 values, $y \in \{0, 1\}$. Assume

$$\mathbb{P}(y = 1|x) = x^T \beta_0$$

(The conditional distribution of y given x is Bernoulli.)

Case 2: y takes positive integer values, $y \in \{0, 1, 2, \dots\}$. Assume

$$\mathbb{P}(y = k|x) = \frac{\exp(-x^T \beta_0)(x^T \beta_0)^k}{k!} \quad k = 0, 1, 2, \dots$$

(The conditional distribution of y given x is Poisson.)

(a) For Case 1, compute $\mathbb{E}[y|x]$. Does this justify a linear regression model of the form $y = x^T \beta_0 + u$?

- Expected value of a Bernoulli distribution is just probability of success p , so $\mathbb{E}[y|x] = p = x^T \beta_0$.

- Can also calculate: $\mathbb{E}[y|x] = 1 * \sum x_i \beta_0 + 0 * (1 - \sum (x_i \beta_0))$ which is equivalent to $x^T \beta_0$.
 - Since $\mathbb{E}[y|x] = x^T \beta_0$, regression model in the form of $y = x^T \beta_0 + u$ is justified to approximate $y = \mathbb{E}[y|x] + u$.
- (b) For Case 1, compute $\text{Var}(y|x)$. Does this justify an alternative estimator to OLS?
- For Bernoulli distribution, variance is $p(1-p)$, so variance here is $\text{Var}(y|x) = x^T \beta_0 (1 - x^T \beta_0)$.
 - Since variance is dependent on x , this case is heteroskedastic. OLS requires an assumption of homoskedasticity, an alternative estimator would be justified.
- (c) For Case 2, compute $\mathbb{E}[y|x]$. Does this justify a linear regression model of the form $y = x^T \beta_0 + u$?
Hint:

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp \lambda$$

- Since this is a Poisson distribution, with $\lambda = x^T \beta_0$, we know the expected value $\mathbb{E}[y|x] = \lambda = x^T \beta_0$. But let's calculate it anyway!

$$\begin{aligned}
 \mathbb{E}[y|x] &= \sum_{k=0}^{\infty} \left(k \times \frac{\exp(-x^T \beta_0) (x^T \beta_0)^k}{k!} \right) && (\mathbb{E} \text{ of discrete dist}) \\
 &= \exp(-x^T \beta_0) \sum_{k=0}^{\infty} \frac{(x^T \beta_0)^k}{(k-1)!} && (\text{algebra}) \\
 &= x^T \beta_0 \exp(-x^T \beta_0) \sum_{k=0}^{\infty} \frac{(x^T \beta_0)^{k-1}}{(k-1)!} && (\text{rearrange}) \\
 &= x^T \beta_0 \exp(-x^T \beta_0) \sum_{j=-1}^{\infty} \frac{(x^T \beta_0)^j}{(j)!} && (\text{define } j = k - 1) \\
 &= x^T \beta_0 \exp(-x^T \beta_0) \sum_{j=0}^{\infty} \frac{(x^T \beta_0)^j}{(j)!} && (\text{assume } \mathbb{P}(y = -1|x) = 0) \\
 &= x^T \beta_0 \exp(-x^T \beta_0) \exp(x^T \beta_0) && (\text{from the hint}) \\
 &= x^T \beta_0 && \blacksquare
 \end{aligned}$$

- Since $\mathbb{E}[y|x] = x^T \beta_0$ (aha, as I suspected!), regression model in the form of $y = x^T \beta_0 + u$ is justified.