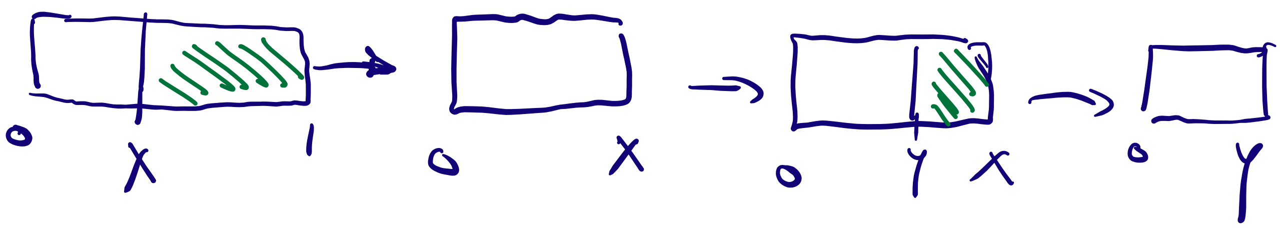
Conditional variance & regression

Broken stick



$X \in U(0,1)$ $\qquad$ $Y \in U(0,X)$

We showed that $\mathbb{E}[Y|X] = \frac{X}{2}$.

Compute variance of the length of the remaining stick,

$$\operatorname{Var} Y = \mathbb{E}(Y - \mathbb{E}(Y))^2 \quad //$$

Def: X, Y have joint distribution. Conditional variance of Y given $X=x$ is

$$\operatorname{Var}(Y|X=x) := \mathbb{E}\Big(\underbrace{\underbrace{(Y - \mathbb{E}[Y|X=x])^2}_{\geq 0}|X=x}_{\geq 0}\Big) \quad //$$

check!

Note that this a function of x, say, $v(x)$.

$$\operatorname{Var}(Y|X) = v(X) \quad \text{random variable.}$$

Exercise: $\operatorname{Var}(Y|X) = \frac{X^2}{12}$ for broken stick. Use variance of $U(0,x)$ is $\frac{x^2}{12}$.

Thm 1: Let X and Y be random variables and g a real-valued function. If $\mathbb{E}Y^2 < \infty$ and $\mathbb{E}(g(X))^2 < \infty$, then

1.) $$\mathbb{E}(Y - g(X))^2 = \mathbb{E}(\mathrm{Var}(Y|X)) + \underbrace{\mathbb{E}\big[\underbrace{(\mathbb{E}(Y|X) - g(X))^2}_{\geq 0}\big]}_{\geq 0}$$
$$\geq \mathbb{E}(\mathrm{Var}(Y|X)).$$

2.) $$\mathrm{Var}\, Y = \mathbb{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbb{E}(Y|X))$$

"law of total variance".

Proof: See AG.

Application of 2): Broken stick: $\mathrm{Var}\, Y = \frac{7}{144}$

$$\mathrm{Var}(Y|X) = \frac{X^2}{12} \Rightarrow \mathbb{E}\,\mathrm{Var}(Y|X) = \frac{\mathbb{E}X^2}{12} \overset{X \in U(0,1)}{=} \frac{\int_0^1 x^2\,dx}{12} = \frac{1}{12}\cdot\frac{1}{3} = \frac{1}{36}.$$

$$\mathrm{Var}\,\mathbb{E}(Y|X) = \mathrm{Var}\,\frac{X}{2} = \frac{1}{4}\mathrm{Var}\,X = \frac{1}{4}\cdot\frac{1}{12} = \frac{1}{48}.$$

According to 2.) above: add the two contributions up

$$\mathrm{Var}\, Y = \frac{1}{36} + \frac{1}{48} = \frac{7}{144}.$$

Upshot: did not use the explicit distribution of Y.

Application of 1): Regression

Setup $X_1, X_2, \ldots, X_n$ and Y are jointly distributed.

"random inputs" "random output"

$$h(x_1, x_2, \ldots, x_n) := \mathbb{E}[Y \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$$
$$= \mathbb{E}[Y \mid \vec{X} = \vec{x}]$$

Choose $n=1$ for simplicity.

- h is called regression function for Y on X
- A predictor for Y based on X is a function $d(X)$.
- Prediction error: $Y - d(X)$, a random variable.
- Expected quadratic prediction error is defined as

$$\mathbb{E}(Y - d(X))^2$$

- If d_1 and d_2 are predictors, then d_1 is better than d_2 if

$$\mathbb{E}(Y - d_1(X))^2 \leq \mathbb{E}(Y - d_2(X))^2.$$

Thm 2: Suppose that $\mathbb{E}\, Y^2 < \infty$. Then the regression function (a.k.a. conditional expectation)

$$h(X) = \mathbb{E}[Y|X] \text{ is}$$

the best predictor for Y based on X. //

Proof: Use Thm 1:

$$\mathbb{E}\,(Y - d(X))^2 = \mathbb{E}\,\mathrm{Var}(Y|X) + \mathbb{E}(\underbrace{\overset{\mathbb{E}[Y|X]}{\overset{\|}{h}}(X) - d(X)}_{=0})^2$$

to be minimized: choose $\boxed{d(X) = h(X).}$

$$\geq \mathbb{E}\,\mathrm{Var}(Y|X)$$

∎

"Among all functions depending on X, $h(X) = \mathbb{E}[Y|X]$ is the best prediction for Y"

Exercise: If X and Y are independent, then

$$\mathrm{Var}(Y|X) = \mathrm{Var}(Y).$$

//