Note 07 - Jan 24

3. Conditional distribution and conditional expectation (cont'd)

3.2. Conditional Expectation (cont'd)

$$\mathbb{E}(g(X)|Y=y) = \begin{cases} \sum_{i_1}^{\infty} g(x_i)P(X=x_u|Y=y) & \text{if } X|Y=y \text{ is discrete} \\ \\ \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)dx & \text{if } X|Y=y \text{ is continuous} \end{cases}$$

3.2.1 What is $\mathbb{E}(X|Y)$?

Different ways to understand *conditional expectation*

1. Fix a value y, $\mathbb{E}(g(X)|Y=y)$ is a number
2. As y changes $\mathbb{E}(g(x)|Y=y)$ becomes a function of y (that each y gives a value): $h(y) =: \mathbb{E}(g(x)|Y=y)$
3. since y is actually random, we can define $\mathbb{E}(g(x)|Y) = h(Y)$. This is a random variable

$$\mathbb{E}(g(x)|Y))_{(\omega)} = \mathbb{E}(g(x)|Y=Y(\omega)$$

$\omega \in \Omega$ this random variable takes value $\mathbb{E}(g(x)|Y=y)$ When $Y=y$

$$\Omega \to \mathbb{R}$$
$$h(Y)_{(\omega)} = h(Y(\omega))$$

3.2.2 Properties of conditional expectation

1. Linearity (inherited from expectation)

$$\mathbb{E}(aX+b|Y=y) = a\mathbb{E}(X|Y=y)+b$$

$$\mathbb{E}(X+Z|Y=y) = \mathbb{E}(X|Y=y)+\mathbb{E}(Z|Y=y)$$

2. $\mathbb{E}(g(X,Y)|Y=y) = \mathbb{E}(g(X,y)|Y=y) \neq \mathbb{E}(g(X,y))$ when X and Y are not independent

Proof (Discrete):

$$\mathbb{E}(g(X,Y)|Y=y) = \sum_{x_i}\sum_{y_j} g(x_i,y_j)\cdot P(X=x_i, Y=y_j|Y=y)$$

$$P(X=x_i,Y=y_j|Y=y) = \begin{cases} 0 & \text{if } y_j \neq y \\ \\ P(X=x_1,Y=y_j)/P(Y+y) = P(X=x_i|Y=y) & \text{if } y_j = y \end{cases}$$

$$\begin{aligned} \Rightarrow \mathbb{E}(g(X,Y)|Y=y) &= \sum_{x_i} g(x_i,y)\cdot P(X=x_i|Y=y) \\ &= \mathbb{E}(g(X,y)|Y=y) & g(X,y) \text{ regarded as a function of } x \end{aligned}$$

In particular,

$$\mathbb{E}(g(X)\cdot h(Y)|Y=y) = h(y)\mathbb{E}(g(X)|Y=y)$$

$$\mathbb{E}(g(X)\cdot h(Y)|Y) = h(Y)\mathbb{E}(g(X)|Y)$$

3. If $X \perp Y$, then $\mathbb{E}(g(X)|Y=y) = \mathbb{E}(g(X))$

Fact: If $X \perp Y$, then conditional distribution of X given $Y=y$ is the same as the unconditional distribution of X

Proof(Discrete):

$$
\begin{aligned}
&\text{if } X \perp Y \\
&P(X = x_i | Y = y_j) \\
&\quad = P(X = x_i | Y = y_j)/P(Y = y_j) \\
&\quad = P(X = x_i)P(Y = y_j)/P(Y = y_j) \\
&\quad = P(X = x_i)
\end{aligned}
$$

4. Law of iterated expectation (or double expectation): Expectation of conditionally expectation is its unconditional expectation

$$\mathbb{E}(\mathbb{E}(X|Y))) = \mathbb{E}(X)$$

$\mathbb{E}(X|Y)$ is a r.v, a function of Y.
Proof(Discrete):

When $Y = y_j$, the r.v. $\mathbb{E}(X|Y) = \mathbb{E}(X|Y = y_j) = \sum_{x_i} x_i P(X = x_i | Y = y_j)$. This happens with probability $P(Y = y_j)$

$$
\Rightarrow \quad
\begin{aligned}
\mathbb{E}(\mathbb{E}(X|Y)) &= \sum_{y_j}(\sum_{x_i} x_i P(X = x_i | Y = y_j))P(Y = y_j) \\
&= \sum_{x_i}\sum_{y_j} P(X = x_i | Y = y_j)P(Y = y_j) \\
&= \sum_{x_i} x_i \sum_{y_j} P(X = x_i | Y = y_j)P(Y = y_2) \qquad \text{law of total probability} \\
&= \sum_{x_i} x_i P(X = x_i) = \mathbb{E}(X)
\end{aligned}
$$

Alternatively,

$$
\begin{aligned}
&\sum_{x_i}\sum_{y_j} x_i P(X = x_i | Y = y_j)P(Y = y_j) \\
&\quad = \sum_{x_i}\sum_{y_j} x_i P(X = x_i, Y = y_j) \qquad\qquad g(X, Y) = X \text{ at } (x_i, y_j) \\
&\quad = \mathbb{E}(X)
\end{aligned}
$$

Example:

Y: # of claims receive by insurance company
X: some random parameter

$$Y|X \sim Poi(X), X \sim Exp(\lambda)$$

a) $\mathbb{E}(Y)$?
b) $P(Y = n)$?

a)

$$Y|X \sim Poi(X) \Rightarrow \mathbb{E}(Y|X = x) = x \Rightarrow \mathbb{E}(Y|X) = X$$

$$
\therefore \begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(\mathbb{E}(Y|X)) \\
&= \mathbb{E}(X) = \frac{1}{\lambda}
\end{aligned}
$$

b)

$$
\begin{aligned}
P(Y=n) &= \int_0^\infty P(Y=n|X=x)f_x(x)dx \\
&= \int_o^\infty \frac{e^{-x}x^n}{n!}\cdot\lambda e^{-\lambda x}dx \\
&= \frac{\lambda}{n!}\int_0^\infty x^n e^{-(\lambda+1)x}dx \\
&= \frac{\lambda}{(\lambda+1)^{n+1}n!}\int_0^\infty ((\lambda+1)x)^n e^{-(\lambda+1)x}d(\lambda+1)x \\
&= \frac{\lambda}{(\lambda+1)^{n+1}n!}\Gamma(n+1) && \Gamma(n+1)=n!\text{ ; formula for gamma function or integration by parts} \\
&= \frac{\lambda}{(\lambda+1)^{n+1}} = (\frac{1}{\lambda+1})^n\cdot\frac{1}{\lambda+1} \\
&\Rightarrow Y+1\sim Geo(\lambda/(\lambda+1))
\end{aligned}
$$

We verify that $\mathbb{E}(Y)=\frac{\lambda+1}{\lambda}-1=\frac{1}{\lambda}$

3.3 Decomposition of variance (EVVE's low)

Definition: The conditional variance of Y given $X=x$ is defined as

$$Var(Y|X=x)=\mathbb{E}((Y-\mathbb{E}(Y|X=x))^2|X=x)$$

$$Var(Y|X)_{(\omega)}=Var(Y|X=X_{(\omega)}) \qquad Var(Y|X)_{(\omega)}\text{: a r.v, a function of }X$$

The conditional variance is simply the variance taken under the conditional distribution

$$\Rightarrow V(Y|X=x)=\mathbb{E}(Y^2|X=x)-(\mathbb{E}(Y|X=x))^2$$

Thus

$$Var(Y)=\mathbb{E}(Var(Y|X))+Var(\mathbb{E}(Y|X))$$

$$\mathbb{E}(Var(Y|X))\text{ : "intra-group variance"} \qquad Var(\mathbb{E}(Y|X))\text{ : "inter-group variance"}$$

Proof:

$$
\begin{aligned}
RHS &= E(E(Y^2|X)-(E(Y|X))^2)+E((E(Y|X))^2)-(E(E(Y|X)))^2 \\
&= E(E(Y^2|X))-\cancel{E((E(Y|X))^2)}+\cancel{E((E(Y|X))^2)}-(E(E(Y|X)))^2 \\
&= E(Y^2)-(E(Y))^2 \\
&= Var(Y)
\end{aligned}
$$