Discrete: $E(Y|A) = \sum_y y \cdot P(Y=y|A)$

Continous: $E(Y|A) = \int_{-\infty}^{\infty} y f(y|A)\,dy$.

Theorem: Law of total expectation.

Let $A_1, A_2, \ldots A_n$ be events which partition the example, with $P(A_i) > 0$, and Y is a r.v. on this sample space.

$$E(Y) = \sum_{i=1}^{n} E(Y|A_i) \times P(A_i)$$

special case: $E(Y) = E(Y|A) \cdot P(A) + E(Y|A^c) \cdot P(A^c)$

Properties:

(1) if X, Y indep $E(Y|X) = E(Y)$

(2) $E(h(X) \cdot Y|X) = h(X)\, E(Y|X)$

(3) $E(Y_1 + Y_2|X) = E(Y_1|X) + E(Y_2|X)$

(Law of Iterated Expeetion Adam's Law)

Theorem: for any r.v.'s X & Y

$E[E(Y|X)] = E(Y)$

Proof: $g(x) = E(Y|X)$

$$Eg(x) = \sum_x g(x) P(X=x) = \sum_x E(Y|X) P(X=x)$$
$$= \sum_x \left(\sum_y y \cdot P(Y=y|X=x) \right) P(X=x)$$
$$= \sum_x \sum_y y\, P(Y=y, X=x)$$
$$= \sum_y y\, P(Y=y)$$
$$= E(Y)$$

Theorem: (Adam's Law with extra conditioning)

$E(E(Y|X,Z)|Z) = E(Y|Z)$

proof: $g(x,z) = E(Y|X,Z)$

$$E(g(x,z)|Z) = \sum_x g(x,z) P(X=x|Z)$$
$$= \sum_x E(Y|X,Z) P(X=x|Z)$$
$$= \sum_x \sum_y y\, P(Y=y|X=x, Z) P(X=x|Z)$$
$$= \sum_x \sum_y y\, P(Y=y, X=x|Z)$$
$$= \sum_y y\, P(Y=y|Z) = E(Y|Z).$$

Theorem: $a = EY$ minimize $E(Y-a)^2$, a is constant.

Proof: $E(Y-a)^2 = \sum_y (y-a)^2 P(Y=y)$

$= \sum_y (y^2+a^2-2ay)P(Y=y)$

$= E(y^2)+a^2-2aE(y)$

$\frac{\partial E(Y-a)^2}{\partial a} = 2a-2E(y) \Rightarrow a = E(y)$ gets minimum.

$(X,Y) \rightarrow$ a random vector.

Observe x, Predict $\hat{y} = h(x) \Rightarrow \min E(Y-h(x))^2$

Theorem: $g(x) = E(Y|X)$ is the best prediction.

Proof: $E(Y-h(x))^2 = E(Y-g(x)+g(x)-h(x))^2$

$= E(Y-g(x))^2 + E(g(x)-h(x))^2$

$+2\boxed{E(Y-g(x))(g(x)-h(x))}$

$E(Y-g(x))(g(x)-h(x))$

$= \underline{E\{E(Y-g(x))(g(x)-h(x))|X\}}$ (by Adam's Law)

$= E\{(g(x)-h(x))\boxed{E(Y-g(x))|}X\}$

$$E(Y-g(x)|x) = E(Y|x) - E(g(x)|x)$$
$$= g(x) - g(x) = 0$$

Hence $E(Y-h(x))^2 = E(Y-g(x))^2 + E(g(x)-h(x))^2$
$$\geq E(Y-g(x))^2$$

$\Longrightarrow$ $g(x)$ is the best prediction.

$y = a + bx + \varepsilon$, $E(\varepsilon|x) = 0 \Rightarrow E\varepsilon = EE\varepsilon|x = 0$

$EY = a + bEX$

$$Cov(x,y) = E(x-EX)(Y-EY) = E(x-Ex)(bx+\varepsilon-bEx)$$
$$= b\,Var\,x + E(x\cdot\varepsilon) - E[(Ex)\cdot\varepsilon]$$
$$= b\,Var\,x + E\,E\,x\cdot\varepsilon|x\,|\,x$$
$$= b\,var\,x$$

$b = \dfrac{Cov(X,Y)}{Var\,X}$ $\quad a = EY - EX\cdot\dfrac{Cov(X,Y)}{Var\,X}$