37161 Probability and Random Variables

Lecture 6

Conditional Expectation

- Consider rolling two regular fair six-sided dice and multiplying the two numbers shown.

- Let X be the number shown on the first die and let Y be the product of the number shown on the two dice.

- We have already seen that the expected value shown when rolling one regular die is 3.5 so, as the numbers shown on the two dice are independent $E(Y) = 3.5 \times 3.5 = 12.25$
- If we have an observation of the value of X, how does this impact our expectation of the value of Y?

Conditional Expectation

- Clearly, if we have observed that $X = 1$, then the product Y is simply equal to 1 multiplied by the number on the second die, so $E(Y \mid X = 1) = 3.5$.

- Similarly, if we have observed that $X = 2$, then the product Y is simply equal to 2 multiplied by the number on the second die, so $E(Y \mid X = 2) = 7$.

Conditional Expectation

- We can obtain the expectation of *Y* from conditional expectations by “averaging out” the conditional expectations over the uncertainties in *X*.

- $E(Y) = E(Y \mid X = 1)P(X = 1) + E(Y \mid X = 2)P(X = 2) + \ldots + E(Y \mid X = 6)P(X = 6)$

- $E(Y) = (3.5)\frac{1}{6} + (7)\frac{1}{6} + (10.5)\frac{1}{6} + (14)\frac{1}{6} + (17.5)\frac{1}{6} + (21)\frac{1}{6} = \frac{73.5}{6} = 12.25$

- This generalises to the law of total expectation.

Law of Total Expectation

- The **law of total expectation** gives that $E(Y) = E(E(Y \mid X))$ where X and Y are events defined on the same sample space.

- Although the justification on the previous slides only shows that this is the case for summing a finite number of discrete probability masses, it also holds more generally, including for continuous distributions.

Conditional Variance

- We can take a similar approach with expected variances.

- Consider $Var(X) = E(X^2) - E(X)^2$.

- We know that this can be considered in terms of expectations of conditional expectations i.e. $Var(X) = E(E(X^2)|Y) - [E(E(X|Y)]^2$

- We can now write this as $Var(X) = E(E(X^2)|Y) - [E(E(X|Y)]^2 + E(E(X|Y)^2) - E(E(X|Y)^2)$

Law of Total Variance

- Gathering the terms $Var(X) = E(E(X^2)|Y) - [E(E(X|Y)]^2 + E(E(X|Y)^2) - E(E(X|Y)^2)$, we obtain that $Var(X) = \left[E(E(X^2)|Y) - E(E(X|Y)^2)\right] + \left[E(E(X|Y)^2) - [E(E(X|Y)]^2\right]$

- This simplifies to $Var(X) = [E(Var(X|Y))] + [Var(E(X|Y))]$.

- This is the **law of total variance**.

- As variances are always non-negative, this also gives that $Var(X) \geq Var(E(X|Y))$. Having some conditional information can neve increase our uncertainty about the value of X.

Example

- Consider a game whereby a player rolls 20 regular fair six-sided dice and observes how many times the dice show a 6. He/she then selects that number of fair coins and flips them. For example, if the 20 dice show four 6s, the player flips four coins.
- The player then flips this (uncertain) number of coins and scores one point for each coin which lands Heads.
- What are the expectation and the variance of the player’s score?

Example

- Let N be the number of 6s shown by the twenty dice. We know that this value would be binomially distributed $N \sim Bin\left(20, \frac{1}{6}\right)$.

- Let X be the number of coins which land Heads. If we knew N (which we don't), then X would also be binomially distributed $X \mid N \sim Bin\left(N, \frac{1}{2}\right)$

- We know that, for $Y \sim Bin(n, p)$, $E(Y) = np$ and $Var(Y) = np(1-p)$.

Example

- Applying the law of total expectation to $X \mid N \sim Bin\left(N, \frac{1}{2}\right)$ therefore gives us that $E(X) = E(E(X \mid N)) = E\left(\frac{1}{2}N\right)$
- As $N \sim Bin\left(20, \frac{1}{6}\right)$ and hence $E(N) = \frac{20}{6}$, we have that $E(X) = \frac{5}{3}$.
- Similarly, $Var(X \mid N) = N\left(\frac{1}{2}\right)\left(1 - \frac{1}{2}\right) = \frac{N}{4}$
- $Var(X) = \left[E(Var(X \mid N))\right] + \left[Var(E(X \mid N))\right] = E\left(\frac{N}{4}\right) + Var\left(\frac{N}{2}\right)$

Example

- $Var(X) = [E(Var(X \mid N))] + [Var(E(X \mid N))] = E\left(\frac{N}{4}\right) + Var\left(\frac{N}{2}\right)$

$$= E\left(\frac{N}{4}\right) + \frac{1}{4}Var(N)$$

$$= \left(\frac{20}{6}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)\left(20 \times \frac{1}{6} \times \frac{5}{6}\right) = \frac{5}{6} + \frac{25}{36} = \frac{55}{36}$$

- Later in this subject, we will see a method for working out the distribution of the sum of a random number of random numbers which could be used to show that, in fact, $X \sim Bin\left(20, \frac{1}{12}\right)$. Note that the result above is consistent with this.

Example

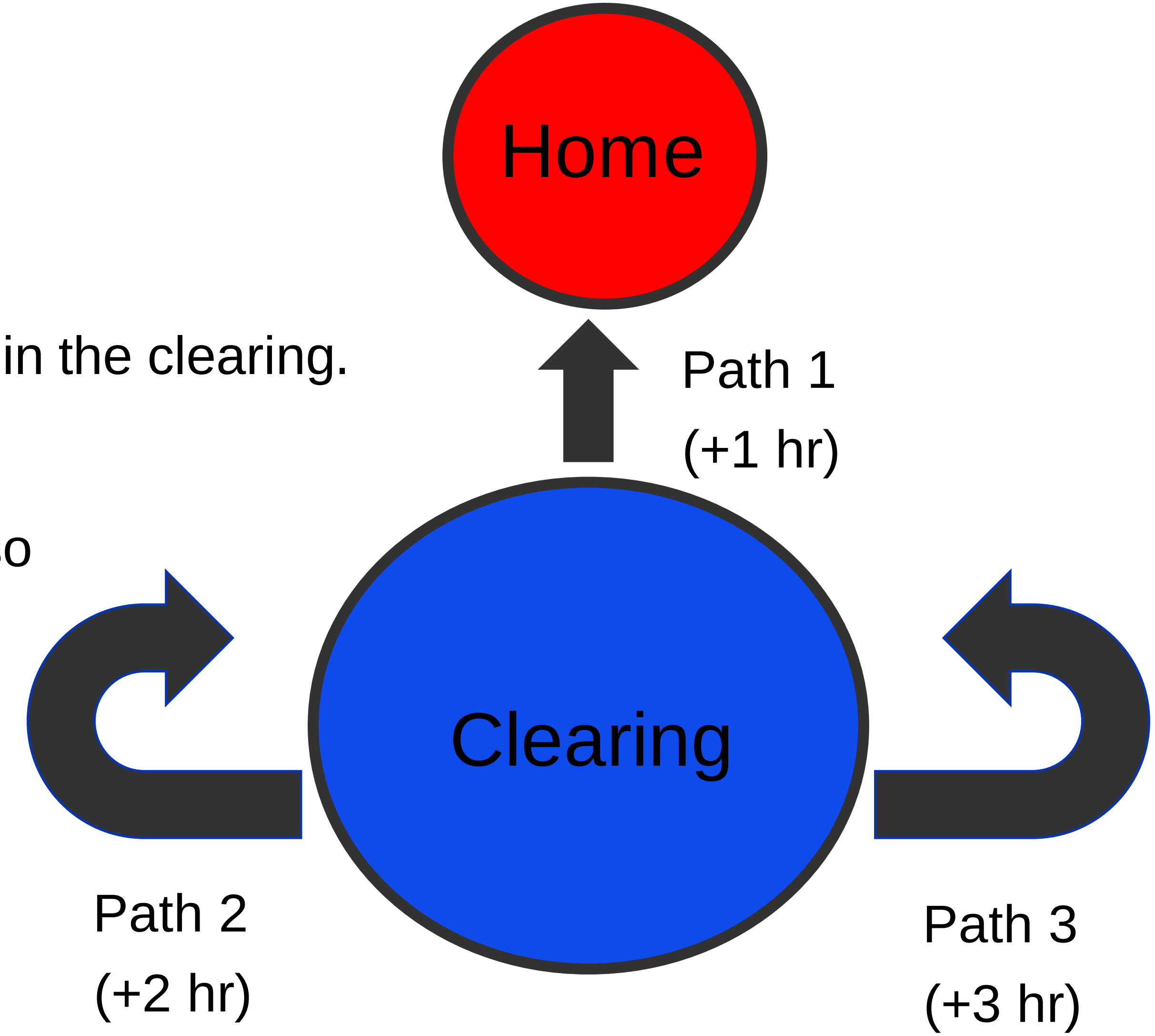
- Consider the case of a walker lost in the woods.
- He/she comes to a clearing in the woods.
- There are three paths:

- If he/she chooses Path 1 he/she will be home in 1 hr.
- If he/she chooses Path 2, he/she will have to turn back and will be back in the clearing in 2 hr.
- If he/she chooses Path 3, he/she will have to turn back and will be back in the clearing in 3hr.

- If each path is chosen with equal probability, what are $E(T)$ and $Var(T)$ if T is the time taken for the walker to get home from the clearing? (Assume that the walker does not remember which paths have previously been taken.)

Example

- Let Y be the number of the path chosen when in the clearing.
- Path 1 takes the walker straight home in 1hr, so $E(T \mid Y = 1) = 1$.
- Path 2 takes 2hr and the walker is again in the clearing, so $E(T \mid Y = 2) = 2 + E(T)$.
- Path 3 takes 3hr and the walker is again in the clearing, so $E(T \mid Y = 3) = 3 + E(T)$.



Example

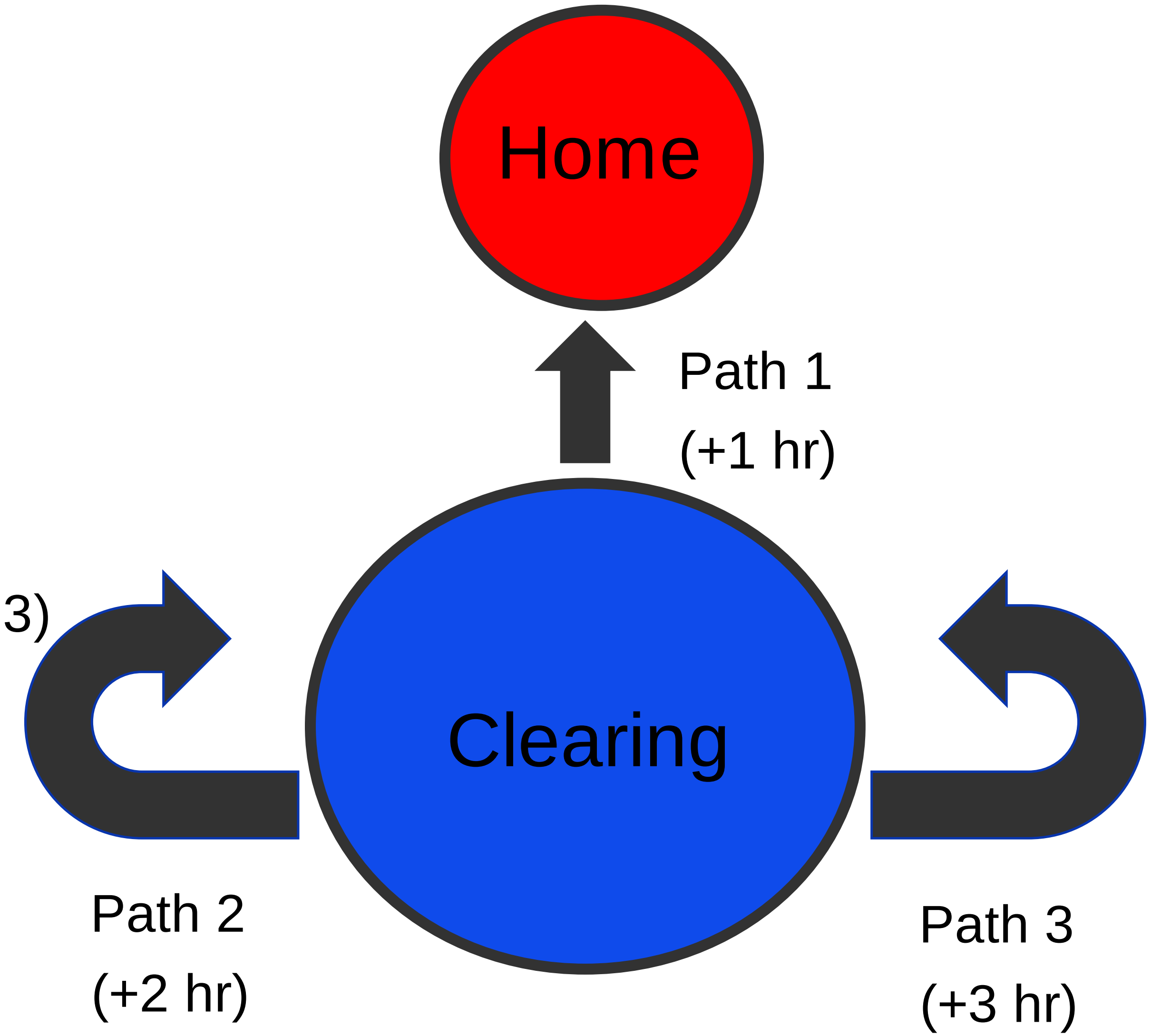
- The law of total expectation tells us that

$$E(T) = E(E(T \mid Y)) = \sum_{k=1}^{3} E(T \mid Y = k) P(Y = k)$$

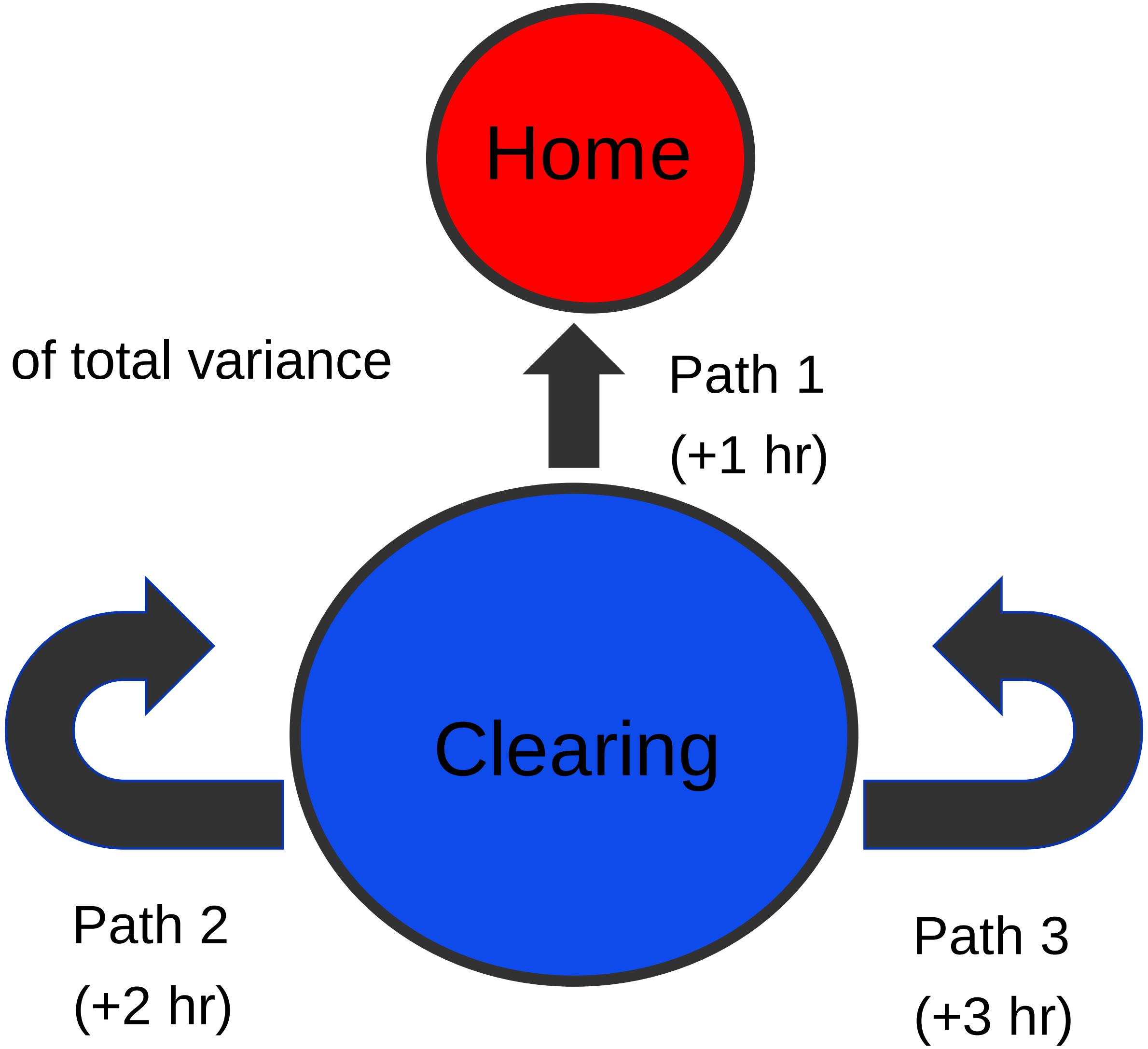
$$= \left(\frac{1}{3}\right) E(T \mid Y = 1) + \left(\frac{1}{3}\right) E(T \mid Y = 2) + \left(\frac{1}{3}\right) E(T \mid Y = 3)$$

$$= \left(\frac{1}{3}\right)(1) + \left(\frac{1}{3}\right)(2 + E(T)) + \left(\frac{1}{3}\right)(3 + E(T))$$

- This gives $E(T) = 2 + \frac{2}{3} E(T)$ hence $E(T) = 6$.



Example



- To calculate the variance of T, we need the law of total variance

$$Var(T) = [E(Var(T \mid Y))] + [Var(E(T \mid Y))]$$

- Taking Path 1 gives no uncertainty about T (T is certainly 1) hence $Var(T \mid Y = 1) = 0$.
- Taking Path 2 only adds 2hr to the time but the walker is again in the clearing hence $Var(T \mid Y = 2) = Var(T + 2) = Var(T)$.
- Taking Path 3 only adds 3hr to the time but the walker is again in the clearing hence $Var(T \mid Y = 3) = Var(T + 3) = Var(T)$.

Example

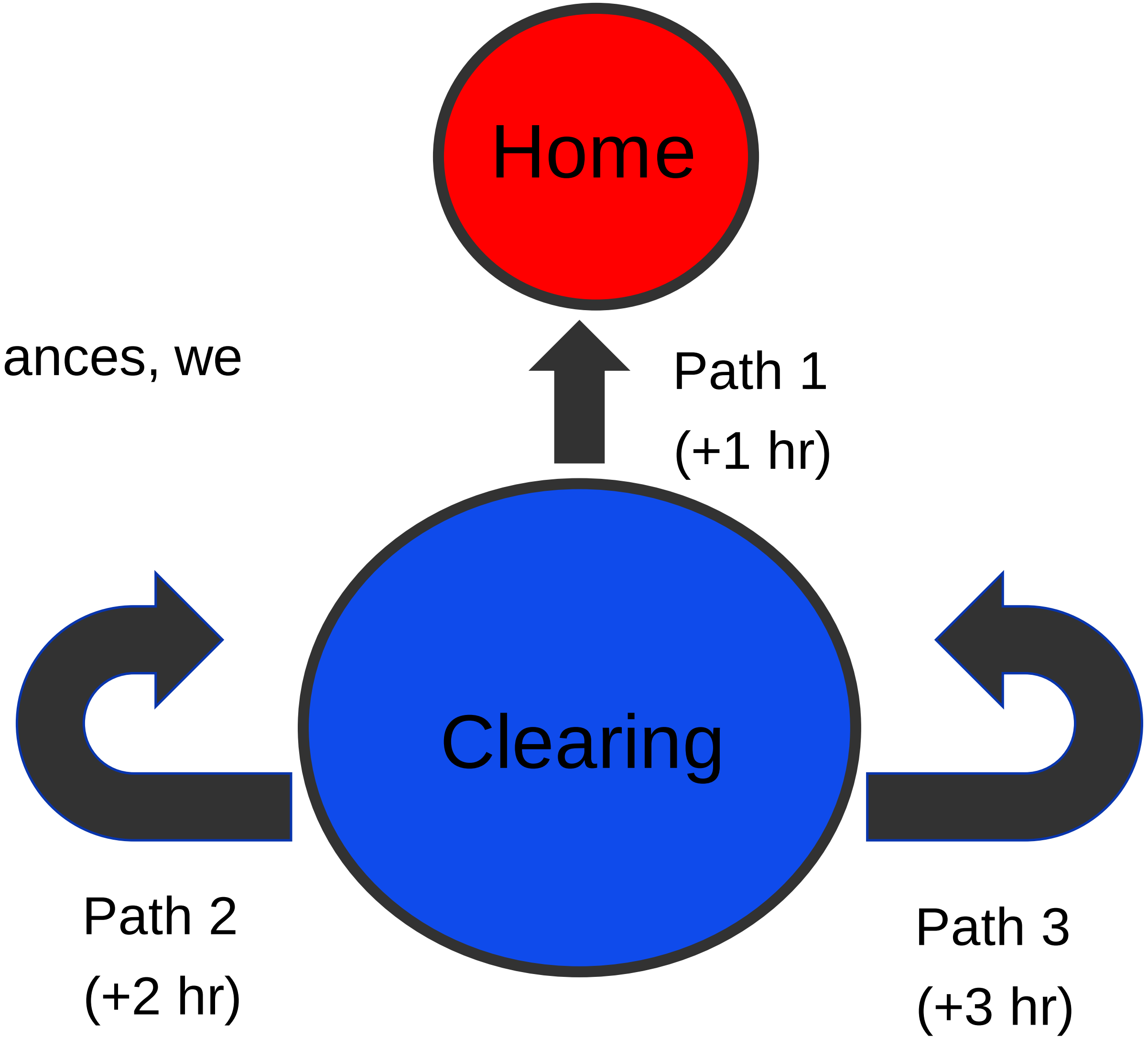
- Taking the expectation of these conditional variances, we obtain $E(Var(T\,|\,Y)) = \sum_{k=1}^{3} Var(T\,|\,Y=k)P(Y=k)$

$$= \left(\frac{1}{3}\right)Var(T\,|\,Y=1) + \left(\frac{1}{3}\right)Var(T\,|\,Y=2)$$

$$+\left(\frac{1}{3}\right)Var(T\,|\,Y=3)$$

$$= \left(\frac{1}{3}\right)(0) + \left(\frac{1}{3}\right)Var(T) + \left(\frac{1}{3}\right)Var(T)$$

$$= \frac{2}{3}Var(T)$$

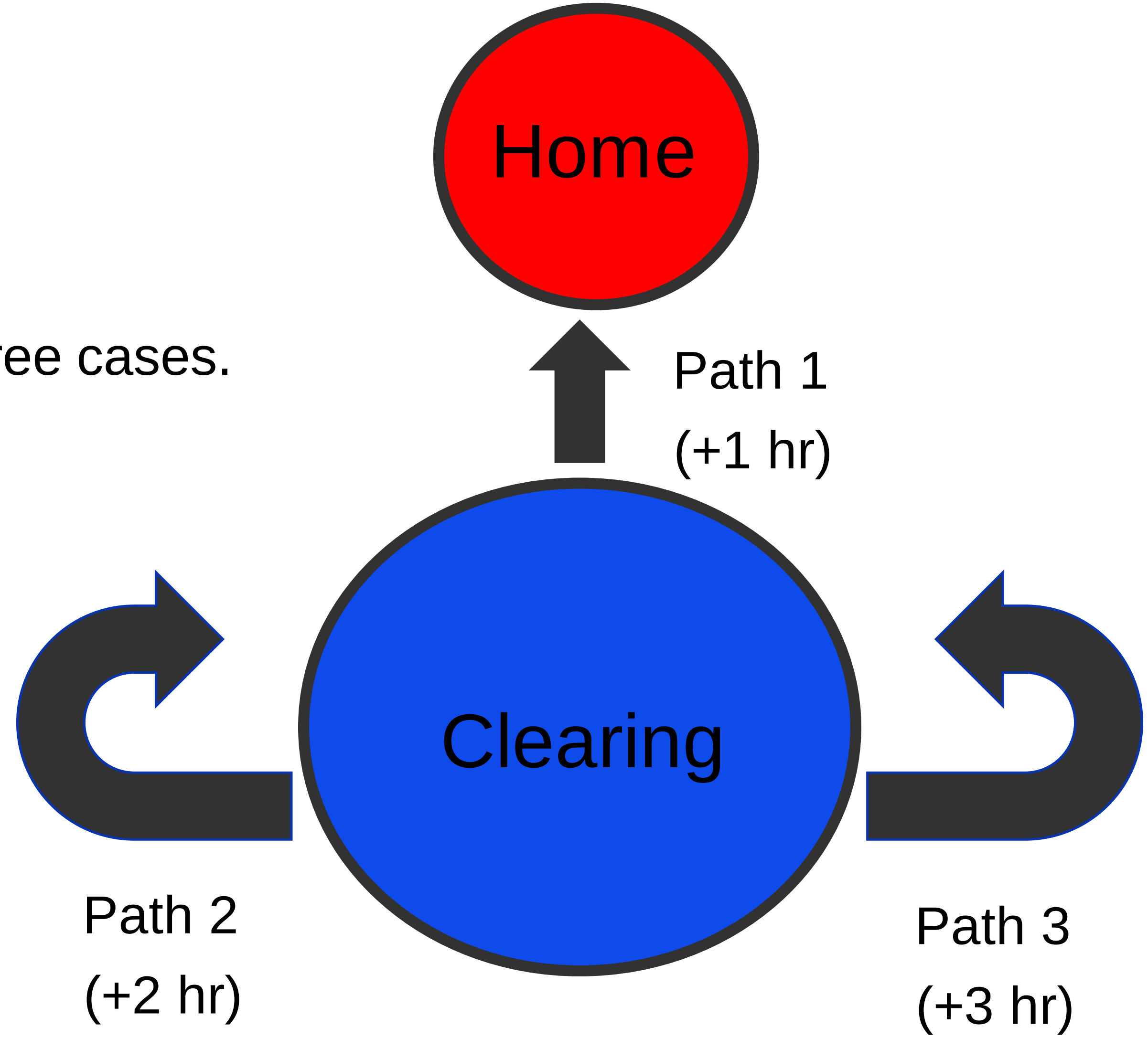


Example

- To calculate $Var(E(T|Y))$, we consider the three cases.

- We know that $E(T|Y=1)=1$
$E(T|Y=2)=2+E(T)$
$E(T|Y=3)=3+E(T)$

- Since $E(T)=6$ and each path is chosen with equal probability, we have a random variable which takes the value 1, 8 or 9, all with probability $\frac{1}{3}$.



Example

- $Var(E(T\mid Y)) = \left(\dfrac{1^2+8^2+9^2}{3}\right) - \left(\dfrac{1+8+9}{3}\right)^2$

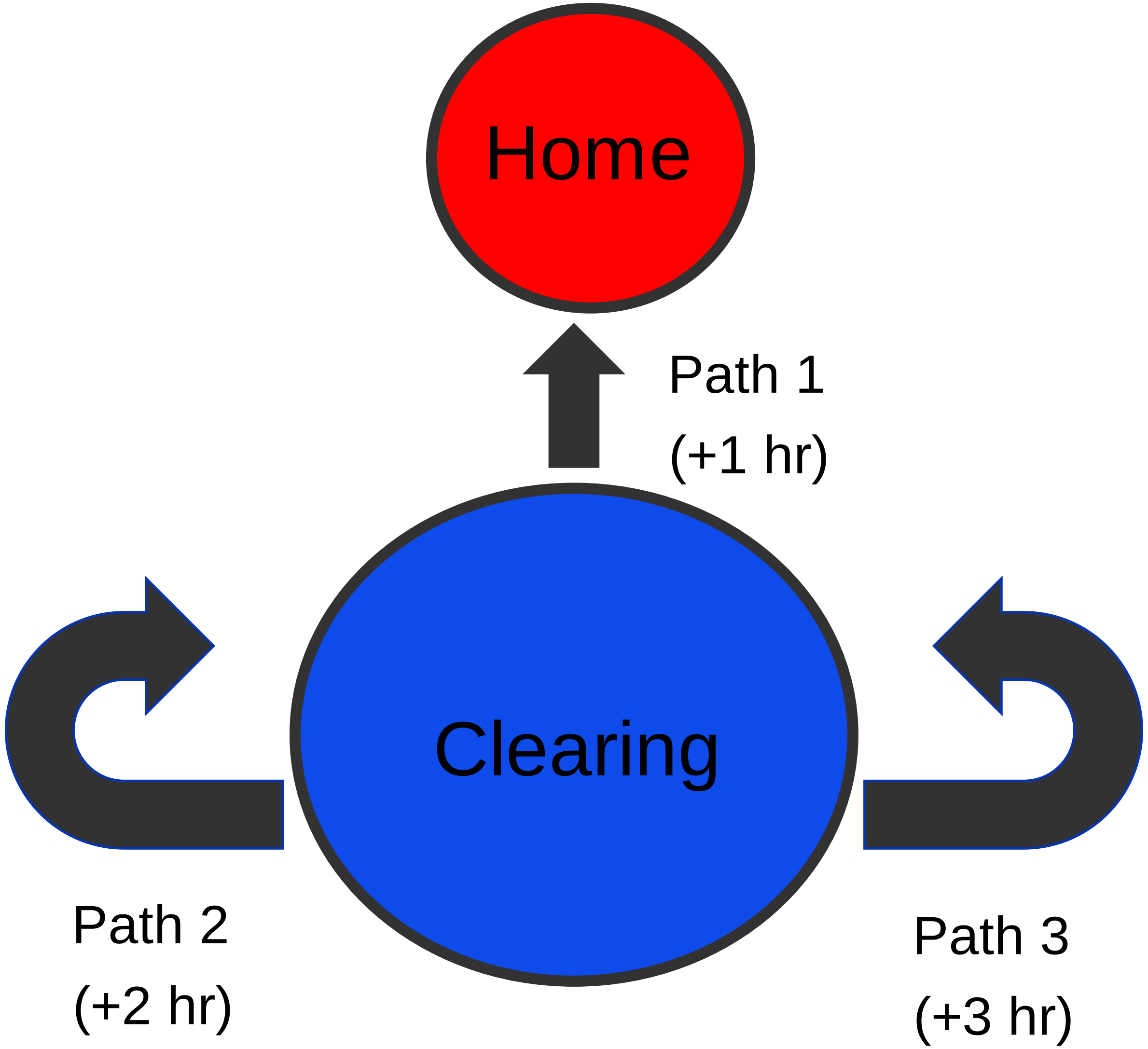
$$= \frac{146}{3} - 36 = \frac{38}{3}.$$

- The law of total variance therefore gives

$$Var(T) = [E(Var(T\mid Y))] + [Var(E(T\mid Y))]$$

$$= \frac{2}{3}Var(T) + \frac{38}{3}.$$

- This gives $Var(T) = 38$.



Distribution of a Function of a Random Variable

- We have already seen that, for the discrete case, obtaining the probability mass function for a function of a discrete random variable is quite straightforward.

- For example, consider X with mass function $P(X=k)=\begin{cases}0.3 & k=1\\0.7 & k=3\\0 & \text{otherwise}\end{cases}$ and the distribution of X^2+1.

- As X is discrete, X^2+1 is also discrete.

- X can only take the values 1 and 3 hence X^2+1 can only take the values 2 and 10.

Distribution of a Function of a Random Variable

- $P(X=k)=\begin{cases} 0.3 & k=1 \\ 0.7 & k=3 \\ 0 & \text{otherwise} \end{cases}$ implies that $P(X^2+1=k)=\begin{cases} 0.3 & k=2 \\ 0.7 & k=10 \\ 0 & \text{otherwise} \end{cases}$

- Note that, when the function transforming the variable is not 1-to-1, we sometimes have to combine masses.

- For example $P(Y=k)=\begin{cases} 0.3 & k=1 \\ 0.3 & k=-1 \\ 0.4 & k=3 \\ 0 & \text{otherwise} \end{cases}$ implies that $P(Y^2+1=k)=\begin{cases} 0.6 & k=2 \\ 0.4 & k=10 \\ 0 & \text{otherwise} \end{cases}$

Distribution of a Function of a Random Variable

- Consider now the continuous case. Let X be a continuous random variable with probability density function $f(x)=\begin{cases}\dfrac{x}{50} & x\in[0,10]\\ 0 & \text{otherwise}\end{cases}$.
- It is easy to verify that this is a valid density function since $\int_0^{10} f(x)\,dx=\int_0^{10}\frac{x}{50}\,dx=\left[\frac{x^2}{100}\right]_0^{10}=1$
- Note, though, that if we want the distribution of a function of X, say, X^2+1, we cannot simply obtain the density function of this by substituting into the density function of X.

Distribution of a Function of a Random Variable

- For $f(x) = \begin{cases} \dfrac{x}{50} & x \in [0,10] \\ 0 & \text{otherwise} \end{cases}$, it is clear that $(X^2+1) \in [1,\ 101]$

- If we let $Y = X^2 + 1$, we <u>cannot</u> simply write $\sqrt{Y-1} = X$ and hence

$$f(y) = \begin{cases} \dfrac{\sqrt{y-1}}{50} & y \in [1,\ 101] \\ 0 & \text{otherwise} \end{cases}$$

- This is not even a valid density function since $\displaystyle\int_1^{101} \frac{\sqrt{y-1}}{50}\,dy = \left[\frac{2}{150}(y-1)^{\frac{3}{2}}\right]_1^{101} = \frac{400}{3} \neq 1.$

Distribution of a Function of a Random Variable

- If we consider integration by substitution, we have that, given a definite integral

$I = \int_a^b f(x)\,dx$ and a continuous differentiable function $y(x)$ then $I = \int_a^b f(x)\,dx = \int_{y(a)}^{y(b)} f(y)\frac{\mathrm{dx}}{dy}\,dy$

- This gives that the density function of $Y(X)$ is given by $g(y) = f(x(y))x'(y)$.

Distribution of a Function of a Random Variable

- Returning to our earlier example with $f(x) = \begin{cases} \frac{x}{50} & x \in [0,10] \\ 0 & \text{otherwise} \end{cases}$ and $Y = X^2 + 1$, we have that

  $\sqrt{Y-1} = X$ and hence $\frac{1}{2\sqrt{Y-1}} = \frac{dX}{dY}$

- This gives the density function of Y as $g(y) = \begin{cases} \left(\frac{\sqrt{y-1}}{50}\right)\left(\frac{1}{2\sqrt{y-1}}\right) & y \in [1,\ 101] \\ 0 & \text{otherwise} \end{cases}$

- Note that $g(y) = \begin{cases} \left(\frac{1}{100}\right) & y \in [1,\ 101] \\ 0 & \text{otherwise} \end{cases}$ integrates to 1 and hence is a valid density function.

Distribution of a Function of a Random Variable

- Consider now the (harder) case of working out the distribution of $Y = X^2 + 1$ where X has density function $f(x) = \begin{cases} \dfrac{3x^2}{2} & x \in [-1,1] \\ 0 & \text{otherwise} \end{cases}$.

- Unlike before, Y is no longer a 1-to-1 function, since X can take positive and negative values and, for example, $Y(0.1) = Y(-0.1) = 1.01$.

- Here, we need to consider the piecewise inverse $X = \begin{cases} -\sqrt{Y-1} & \text{if } X < 0 \\ \sqrt{Y-1} & \text{if } X \geq 0 \end{cases}$

Distribution of a Function of a Random Variable

- We now consider the cumulative density $G(Y) = P(Y \le y) = P(X^2 + 1 \le y)$

$$= P(-\sqrt{y-1} \le X \le \sqrt{y-1})$$

$$= P(X \le \sqrt{y-1}) - P(X \le -\sqrt{y-1})$$

- Differentiating gives $G' = g(y) = f(\sqrt{y-1})\frac{d}{dy}(\sqrt{y-1}) - f(-\sqrt{y-1})\frac{d}{dy}(-\sqrt{y-1})$

- $g(y) = 2f(\sqrt{y-1})\frac{1}{2\sqrt{y-1}}$ or, in full, $g(y) = \begin{cases} 3(y-1)\frac{1}{2\sqrt{y-1}} & y \in [1,2] \\ 0 & \text{otherwise} \end{cases}$

Distribution of a Function of a Random Variable

- $$g(y) = \begin{cases} 3(y-1)\dfrac{1}{2\sqrt{y-1}} & y \in [1,2] \\ 0 & \text{otherwise} \end{cases} \quad \text{so} \quad g(y) = \begin{cases} \dfrac{3\sqrt{y-1}}{2} & y \in [1,2] \\ 0 & \text{otherwise} \end{cases}$$

- Note also that this is a valid density function, since $\int_1^2 \frac{3\sqrt{y-1}}{2}\, dy = 1$