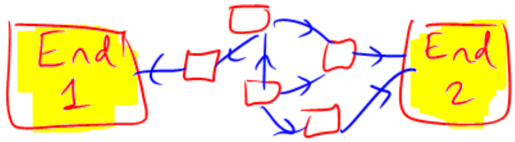
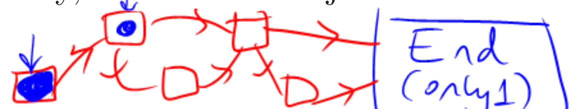
Ch 2: Probability



$\mathbb{P}$(finish in End 1) (not in End 2)

Chapter 3: Expectation and Variance

In the previous chapter we looked at probability, with three major themes:



Ch3

1. Conditional probability: $\mathbb{P}(A \mid B)$.
2. First-step analysis for calculating eventual probabilities in a stochastic process.

   $\mathbb{E}$(# steps to finish)
3. Calculating probabilities for continuous and discrete random variables.

If 2 endings, $\mathbb{E}$(# steps to finish at either end) OK

In this chapter, we look at the same themes for **expectation** and **variance**. The expectation of a random variable is the

But $\mathbb{E}$(# steps to reach End 2) $= \infty$

Imagine observing many thousands of independent random values from the random variable of interest. Take the average of these random values. The expectation is the value of this average as the sample size tends to infinity.

We will repeat the three themes of the previous chapter, but in a different order.

1. Calculating expectations for continuous and discrete random variables.
2. Conditional expectation: the expectation of a random variable X, *conditional* on the value taken by another random variable Y. If the value of Y affects the value of X (i.e. X and Y are *dependent*), the conditional expectation of X given the value of Y will be different from the overall expectation of X.
3. First-step analysis for calculating the expected amount of time needed to reach a particular state in a process (e.g. the expected number of shots before we win a game of tennis).

We will also study similar themes for variance.

$$\mathbb{E}(X^2) = \frac{3^2+3^2+5^2+5^2+5^2+5^2+6^2}{7} = \frac{2*3^2+4*5^2+1*6^2}{7} = \frac{2}{7}*3^2+\frac{4}{7}*5^2+\frac{1}{7}*6^2$$

3.1 Expectation

The **mean**, **expected value**, or **expectation** of a random variable X is written as $\mathbb{E}(X)$ or μ_X. If we observe N random values of X, then the mean of the N values will be approximately equal to $\mathbb{E}(X)$ for large N. The expectation is defined differently for continuous and discrete random variables.

Definition: Let X be a **continuous** random variable with p.d.f. $f_X(x)$. The expected value of X is

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$\mathbb{E}X$ is a real number

Definition: Let X be a **discrete** random variable with probability function $f_X(x)$. The expected value of X is

$$\mathbb{E}(X) = \sum_x x f_X(x) = \sum_x x\,\mathbb{P}(X=x)$$

same again

Expectation of $g(X)$

Let $g(X)$ be a function of X. We can imagine a long-term average of $g(X)$ just as we can imagine a long-term average of X. This average is written as $\mathbb{E}(g(X))$. Imagine observing X many times (N times) to give results $x_1, x_2, \ldots, x_N$. Apply the function g to each of these observations, to give $g(x_1), \ldots, g(x_N)$. The mean of $g(x_1), g(x_2), \ldots, g(x_N)$ approaches $\mathbb{E}(g(X))$ as the number of observations N tends to infinity.

Definition: Let X be a **continuous** random variable, and let g be a function. The expected value of $g(X)$ is

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx.$$

Definition: Let X be a **discrete** random variable, and let g be a function. The expected value of $g(X)$ is

$$\mathbb{E}(g(X)) = \sum_x g(x) f_X(x) = \sum_x g(x)\,\mathbb{P}(X=x)$$

Expectation of XY: the definition of $\mathbb{E}(XY)$

Suppose we have two random variables, X and Y. These might be independent, in which case the value of X has no effect on the value of Y. Alternatively, X and Y might be *dependent*: when we observe a random value for X, it might influence the random values of Y that we are most likely to observe. For example, X might be the height of a randomly selected person, and Y might be the weight. On the whole, larger values of X will be associated with larger values of Y.

person 1 H & W

person 2 H & W

To understand what $\mathbb{E}(XY)$ means, think of observing a large number of *pairs* $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$. If X and Y are dependent, the value x_i might affect the value y_i, and vice versa, so we have to keep the observations together in their pairings. As the number of pairs N tends to infinity, the average $\frac{1}{N}\sum_{i=1}^{N} x_i \times y_i$ approaches the expectation $\mathbb{E}(XY)$.

For example, if X is height and Y is weight, $\mathbb{E}(XY)$ is the average of (height $\times$ weight). We are interested in $\mathbb{E}(XY)$ because it is used for calculating the *covariance* and *correlation*, which are measures of how closely related X and Y are (see Section 3.2).

Properties of Expectation

i) Let g and h be functions, and let a and b be constants. For any random variable X (discrete or continuous),

$$\mathbb{E}\Big\{ag(X) + bh(X)\Big\} = a\mathbb{E}\Big\{g(X)\Big\} + b\mathbb{E}\Big\{h(X)\Big\}.$$

In particular,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$

ii) Let X and Y be ANY random variables (discrete, continuous, independent, or non-independent). Then

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

More generally, for ANY random variables $X_1, \ldots, X_n$,

$$\mathbb{E}(X_1 + \ldots + X_n) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n).$$

iii) Let X and Y be **independent** random variables, and g, h be functions. Then

$$\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y) \text{ only when } X \text{ and } Y \text{ are indept.}$$

$$\mathbb{E}\{g(X)h(Y)\} = \{\mathbb{E}(g(X))\}\{\mathbb{E}(h(Y))\} \quad \text{'' '' '' indept.}$$

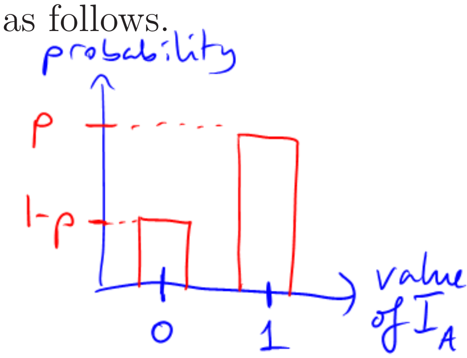
Notes: 1. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ is ONLY generally true if X and Y are INDEPENDENT.

2. If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. However, the converse is not generally true: it is possible for $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ even though X and Y are dependent.

Probability as an Expectation

Let A be any event. We can write $\mathbb{P}(A)$ as an expectation, as follows. Define the ***indicator function:***

$$I_A = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{if } A \text{ doesn't occur.} \end{cases}$$



Then I_A is a ***random variable***, and

$$\mathbb{E}(I_A) = \sum_{r=0}^{1} r\,\mathbb{P}(I_A = r)$$

$$= 0 * \mathbb{P}(I_A = 0) + 1 * \mathbb{P}(I_A = 1)$$

$$= \mathbb{P}(I_A = 1)$$

$$= \mathbb{P}(A \text{ occurs})$$

$$= \mathbb{P}(A)$$

Thus

$$\boxed{\mathbb{P}(A) = \mathbb{E}(I_A)} \quad \text{for any event } A.$$

3.2 Variance, covariance, and correlation

The variance of a random variable X is a measure of how spread out it is. Are the values of X clustered tightly around their mean, or can we commonly observe values of X a long way from the mean value? The *variance* measures how far the values of X are from their mean, on average.

Definition: Let X be any random variable. The **variance** of X is

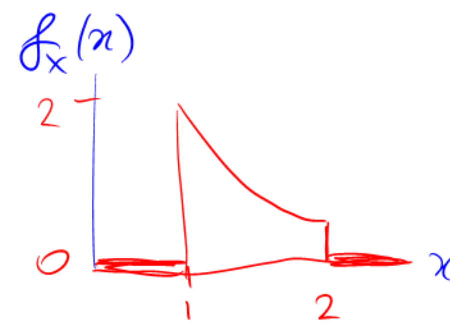
$$\mathrm{Var}(X) = \mathbb{E}\left\{ (X - \mu_X)^2 \right\} = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

The variance is the mean squared deviation of a random variable from its own mean.

If X has *high variance*, we can observe values of X a long way from the mean.

If X has *low variance*, the values of X tend to be clustered tightly around the mean value.

Example: Let X be a continuous random variable with p.d.f.



$$f_X(x) = \begin{cases} 2x^{-2} & \text{for } 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

Find $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_1^2 x \cdot 2x^{-2}\, dx$$

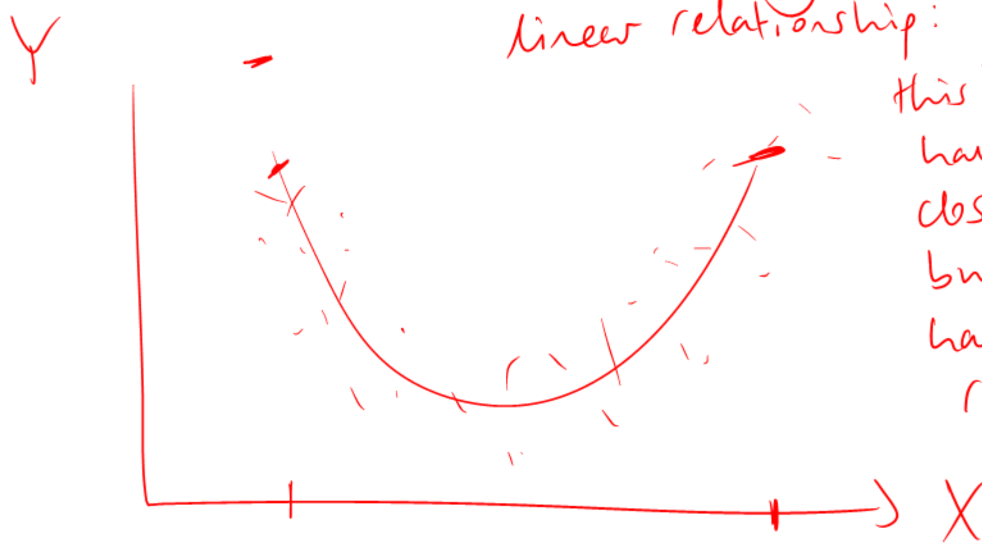
$$= \int_1^2 2x^{-1}\, dx$$

$$= \Big[2 \log x \Big]_1^2$$

$$= 2\log 2 - 2\log 1$$

$$= \underline{2 \log 2}$$

covariance only measures linear relationship:



this X & Y may have covariance close to 0 but they do have a strong relationship

it's just not a linear relationship.

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$
$$= \mathbb{E}(X * X) - (\mathbb{E}X) * (\mathbb{E}X)$$

$$cov(X,Y) = \mathbb{E}(X * Y) - (\mathbb{E}X) * (\mathbb{E}Y)$$

$$cov(X,Y) = \mathbb{E}\{(X - \mu_X) \times (Y - \mu_Y)\}.$$

X & Y are positively related: big X ⟷ big Y
small X ⟷ small Y
positive covariance

X & Y are negatively related: big X ⟷ small Y
small X ⟷ big Y
negative covariance.

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$
$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - (2\log 2)^2$$
$$= \int_1^2 x^2 \cdot 2x^{-2}\,dx - 4(\log 2)^2$$
$$= \int_1^2 2\,dx - 4(\log 2)^2$$
$$= [2x]_1^2 - 4(\log 2)^2$$
$$= 4 - 2 - 4(\log 2)^2$$
$$= 0.0782$$

$\boxed{\log = \ln}$ $\boxed{\log_e}$

Covariance

Covariance is a measure of the association or dependence between two random variables X and Y. Covariance can be either positive or negative. (*Variance* is always positive.)

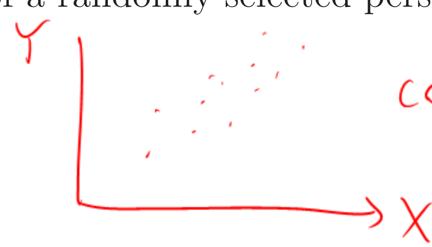
Definition: Let X and Y be any random variables. The **covariance** between X and Y is given by

$$\mathrm{cov}(X,Y) = \mathbb{E}\{(X-\mu_X)(Y-\mu_Y)\} = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$$

where $\mu_X = \mathbb{E}X$ and $\mu_Y = \mathbb{E}Y$

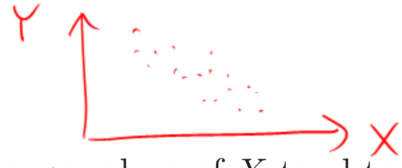
mean of the product minus the product of the means.

1. $\mathrm{cov}(X,Y)$ will be positive if large values of X tend to occur with large values of Y, and small values of X tend to occur with small values of Y. For example, if X is height and Y is weight of a randomly selected person, we would expect $\mathrm{cov}(X,Y)$ to be positive.



$\mathrm{cov}(X,Y) > 0.$

$\text{cov}(X,Y)<0$



2. $\text{cov}(X, Y)$ will be negative if large values of X tend to occur with small values of Y, and small values of X tend to occur with large values of Y. For example, if X is age of a randomly selected person, and Y is heart rate, we would expect X and Y to be negatively correlated (older people have slower heart rates).

3. If X and Y are independent, then there is no pattern between large values of X and large values of Y, so $\text{cov}(X, Y) = 0$. However, $\text{cov}(X, Y) = 0$ does NOT imply that X and Y are independent, unless X and Y are Normally distributed.

Could have a strong non-linear relationship.

indept $X,Y \Rightarrow \text{cov}=0$

$\text{cov}=0 \not\Rightarrow$ indep.

Properties of Variance

i) Let g be a function, and let a and b be constants. For any random variable X (discrete or continuous),

$$\text{Var}\{a\,g(X)+b\} = a^2\,\text{Var}(g(X)).$$

In particular, $\text{Var}(aX+b) = a^2\,\text{Var}(X)$

ii) Let X and Y be **independent** random variables. Then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \qquad \text{indept}$$

iii) If X and Y are **NOT independent**, then

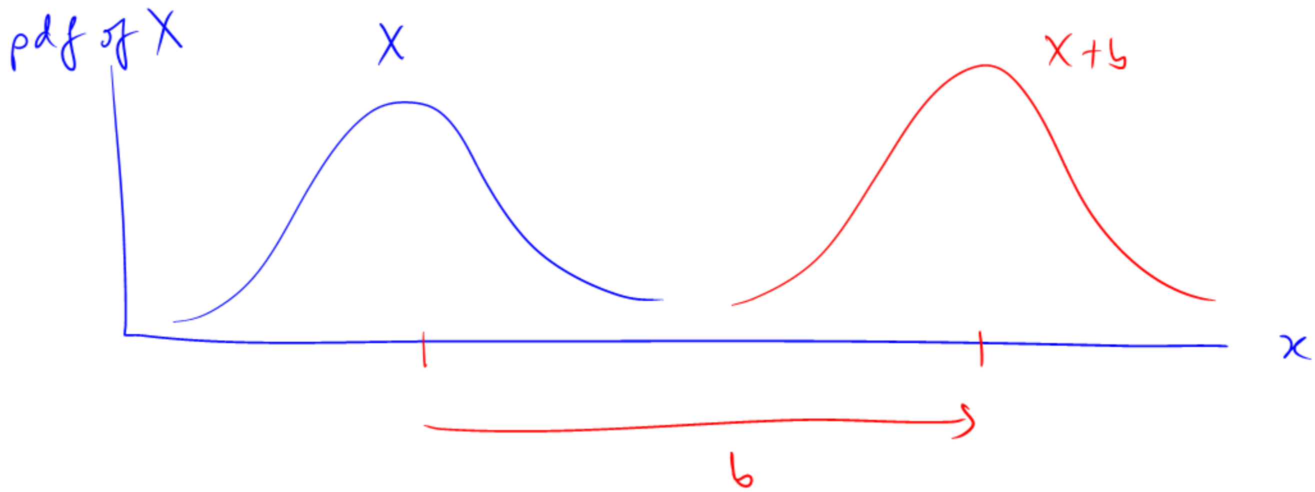
$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{cov}(X,Y).$$

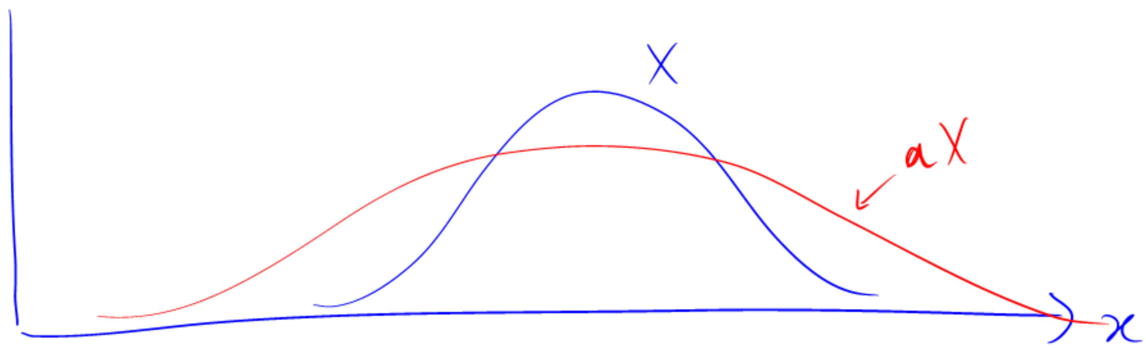
Correlation (non-examinable)

The correlation coefficient of X and Y is a measure of the linear association between X and Y. It is given by the covariance, scaled by the overall variability in X and Y. As a result, the correlation coefficient is always between -1 and $+1$, so it is easily compared for different quantities.

Definition: The **correlation** between X and Y, also called the **correlation coefficient**, is given by

$$\boxed{\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.}$$





$$Y = -X$$

$\mathrm{Var}(X+Y) = \mathrm{Var}(X-X) = \mathrm{Var}(0) = 0$

$\mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$ when non-ind.

The correlation measures <u>linear</u> association between X and Y. It takes values only between -1 and $+1$, and has the same sign as the covariance.

The correlation is ± 1 if and only if there is a perfect linear relationship between X and Y, i.e. $\text{corr}(X, Y) = 1 \iff Y = aX + b$ for some constants a and b.

The correlation is 0 if X and Y are independent, but a correlation of 0 does not *imply* that X and Y are independent.

3.3 Conditional Expectation and Conditional Variance

Throughout this section, we will assume for simplicity that X and Y are discrete random variables. However, exactly the same results hold for continuous random variables too.

Y = height
X = weight

Suppose that X and Y are discrete random variables, possibly dependent on each other. Suppose that we fix Y at the value y. This gives us a set of conditional probabilities $\mathbb{P}(X = x \mid Y = y)$ for all possible values x of X. This is called the conditional distribution of X, given $Y = y$.

Definition: Let X and Y be discrete random variables. The **<u>conditional probability function</u>** of X, given that $Y = y$, is:

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x \text{ AND } Y = y)}{\mathbb{P}(Y = y)}$$

like $\left[\dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\right]$

We write the conditional probability function as:

$$f_{X|Y}(x \mid y) = \mathbb{P}(X = x \mid Y = y).$$

Note: The conditional probabilities $f_{X \mid Y}(x \mid y)$ sum to one, just like any other probability function:

$$\sum_x \mathbb{P}(X = x \mid Y = y) = \sum_x \mathbb{P}_{\{Y=y\}}(X = x) = 1,$$

using the subscript notation $\mathbb{P}_{\{Y=y\}}$ of Section 2.3.

721: use 721 box for 325 assignments.

X = weight, Y = height

We can also find the expectation and variance of X with respect to this conditional distribution. That is, if we know that the value of Y is fixed at y, then we can find the mean value of X *given that* Y takes the value y, and also the variance of X given that $Y = y$. e.g. $\mathbb{E}(X \mid Y = 165\text{cm})$ is the average weight among people of height 165cm.

Definition: Let X and Y be discrete random variables. The **conditional expectation of X, given that $Y = y$**, is

$$\boxed{\mu_{X \mid Y=y} = \mathbb{E}(X \mid Y = y) = \sum_x x f_{X\mid Y}(x \mid y)}$$

$$\boxed{= \sum_x x \, \mathbb{P}(X = x \mid Y = y)}$$

$\mathbb{E}(X \mid Y = y)$ is the mean value of X, when Y is fixed at y.

Conditional expectation as a random variable

The unconditional expectation of X, $\mathbb{E}(X)$, is just a number:
e.g. $\mathbb{E}X = 60$ or $\mathbb{E}X = 72.1$.

The conditional expectation, $\mathbb{E}(X \mid Y = y)$, is a number depending on y.

If Y has an influence on the value of X, then Y will have an influence on the *average* value of X. So, for example, we would expect ~~$\mathbb{E}(X \mid Y = 2)$~~ to be different from ~~$\mathbb{E}(X \mid Y = 3)$~~. $\mathbb{E}(\text{weight} \mid \text{height} = 150\text{cm})$, $\mathbb{E}(\text{weight} \mid \text{height} = 180\text{cm})$

We can therefore view $\mathbb{E}(X \mid Y = y)$ as a function of y, say $\mathbb{E}(X \mid Y = y) = h(y)$.

To evaluate this function, $h(y) = \mathbb{E}(X \mid Y = y)$, we:

i) fix Y at the chosen value y,

ii) find the expectation of X when Y is fixed at this value.

$h(y)$

$h(Y)$: random variable

However, we could also evaluate the function at a *random value* of Y:

i) observe a random value of Y,

ii) fix Y at that observed random value,

iii) evaluate $\mathbb{E}(X \mid Y = \text{observed random value})$.

We obtain a random variable: $\mathbb{E}(X \mid Y) = h(Y)$.

The randomness comes from the randomness in Y, not in X.

Conditional expectation, $\mathbb{E}(X \mid Y)$ is a random variable with randomness inherited from Y, not X.

~~Y = height~~
~~1 m w.p. 1/8~~
~~2 m w.p. 7/8~~

Example: Suppose $Y = \begin{cases} 1 & \text{with probability } 1/8\,, \\ 2 & \text{with probability } 7/8\,, \end{cases}$

and $X \mid Y = \begin{cases} 2Y & \text{with probability } 3/4\,, \\ 3Y & \text{with probability } 1/4\,. \end{cases}$

$$(X \mid Y=2) = \begin{cases} 4 & \text{w.p. } 3/4 \\ 6 & \text{w.p. } 1/4 \end{cases}$$

Conditional expectation of X given $Y = y$ is a number depending on y:

If $Y = 1$, then $X \mid (Y=1) = \begin{cases} 2 & \text{w.p. } 3/4 \\ 3 & \text{w.p. } 1/4 \end{cases}$

So $\mathbb{E}(X \mid Y=1) = 2 * \frac{3}{4} + 3 * \frac{1}{4} = \frac{9}{4}$.

If $Y = 2$, then $X \mid (Y=2) = \begin{cases} 4 & \text{w.p. } 3/4 \\ 6 & \text{w.p. } 1/4 \end{cases}$

So $\mathbb{E}(X \mid Y=2) = 4 * \frac{3}{4} + 6 * \frac{1}{4} = \frac{18}{4}$

So $\mathbb{E}(X \mid Y=y) = \begin{cases} \frac{9}{4} & \text{if } y=1, \\ \frac{18}{4} & \text{if } y=2. \end{cases}$

So $\mathbb{E}(X \mid Y=y)$ is a number depending on y, i.e. a FUNCTION of y.

Conditional expectation of X given random Y is a random variable:

From above, $\mathbb{E}(X \mid Y) = \begin{cases} 9/4 & \text{if } Y=1 \quad (\text{probability } \frac{1}{8}) \\ 18/4 & \text{" } Y=2 \quad (\text{" } \frac{7}{8}) \end{cases}$

So $\mathbb{E}(X \mid Y) = \begin{cases} 9/4 & \text{with probability } 1/8 \\ 18/4 & \text{" " } 7/8 \end{cases}$ ← i.e. it's Y's probabilities that we see here, not X's.

So $\mathbb{E}(X \mid Y)$ is a random variable.
The randomness in $\mathbb{E}(X \mid Y)$ is inherited from Y, not X.

Conditional expectation is a very useful tool for finding the ***unconditional*** expectation of X (see below). Just like the Partition Theorem, it is useful because it is often easier to specify conditional probabilities than to specify overall probabilities.

Conditional variance

The conditional variance is similar to the conditional expectation.

- $\text{Var}(X \mid Y = y)$ is the variance of X, when Y is fixed at the value $Y = y$.
- $\text{Var}(X \mid Y)$ is a random variable, giving the variance of X when Y is fixed at a value to be selected randomly.

Definition: Let X and Y be random variables. The **conditional variance of X, given Y,** is given by

$$\boxed{\begin{aligned} \text{Var}(X \mid Y) &= \mathbb{E}(X^2 \mid Y) - \{\mathbb{E}(X \mid Y)\}^2 \\ &= \mathbb{E}\{(X-\mu_X)^2 \mid Y\} \end{aligned}}$$

Like $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}((X-\mu_X)^2)$ but now all $\mathbb{E}$'s are conditional: $\mathbb{E}(\cdot \mid Y)$.

Like expectation, $\text{Var}(X \mid Y=y)$ is a number depending on y (a function of y), while $\text{Var}(X \mid Y)$ is a random variable with randomness inherited from Y.

Partition Thm = "Law of Total Probability".

Laws of Total Expectation and Variance

If all the expectations below are finite, then for ANY random variables X and Y, we have:

LEARN

Law of Total Expectation

i) $$\boxed{\mathbb{E}(X) = \mathbb{E}_Y\left(\mathbb{E}(X \mid Y)\right)}$$

Note we can pick ANY Y, to make the calculation as easy as we can!

total expectation

$\mathbb{E}_Y$ is new notation, meaning "expectation over Y" or "expectation with respect to the probs of Y"

ii) $$\mathbb{E}(g(X)) = \mathbb{E}_Y\{\mathbb{E}(g(X) \mid Y)\}$$ for any function g.

iii) Law of Total Variance:

$$\boxed{\mathrm{Var}(X) = \mathbb{E}_Y\left(\mathrm{Var}(X \mid Y)\right) + \mathrm{Var}_Y\left(\mathbb{E}(X \mid Y)\right)}$$

MAYBE

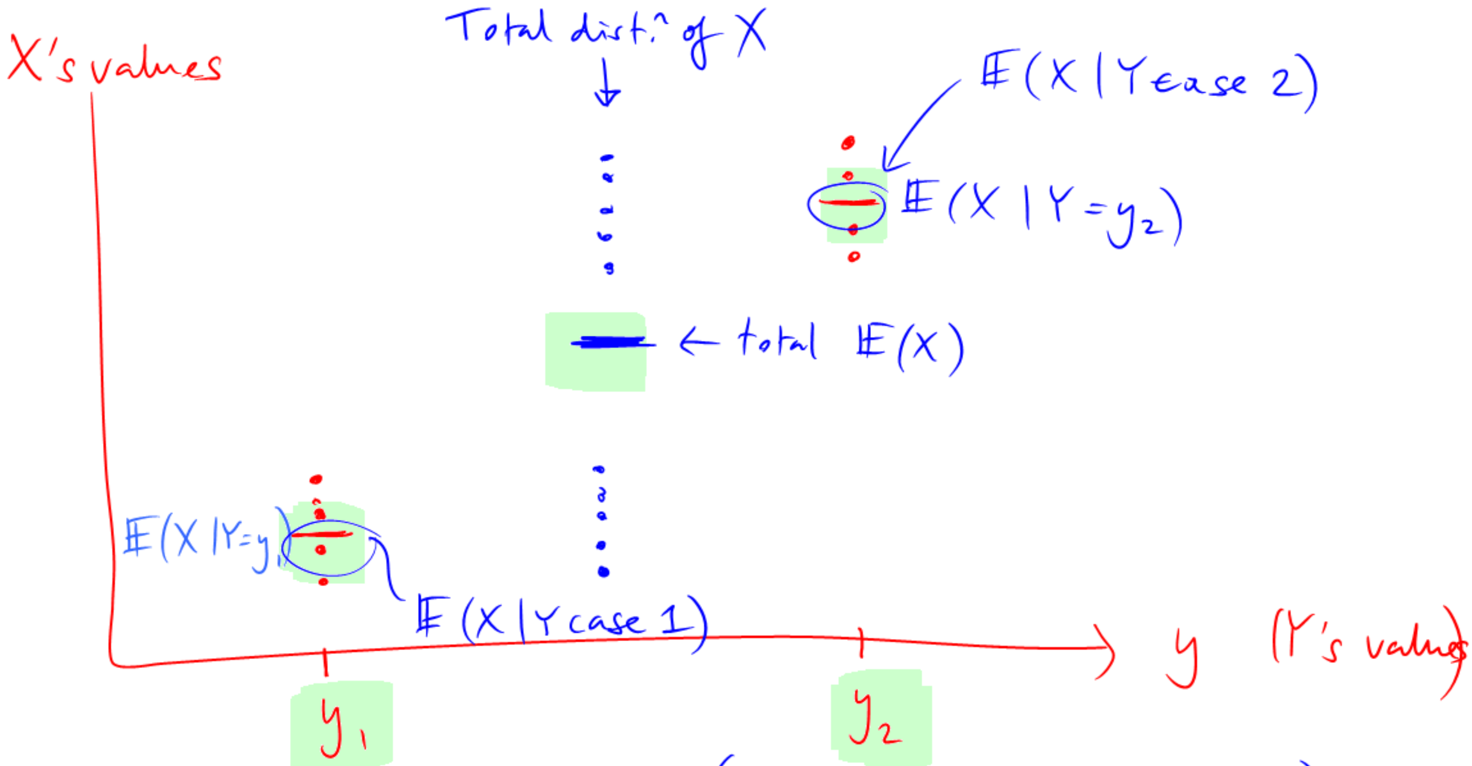
Note: $\mathbb{E}_Y$ and Var_Y denote expectation over Y and variance over Y,

i.e. the expectation or variance is computed over the distribution of the random variable Y.

The Law of Total Expectation says that the total average is the average of case-by-case averages.

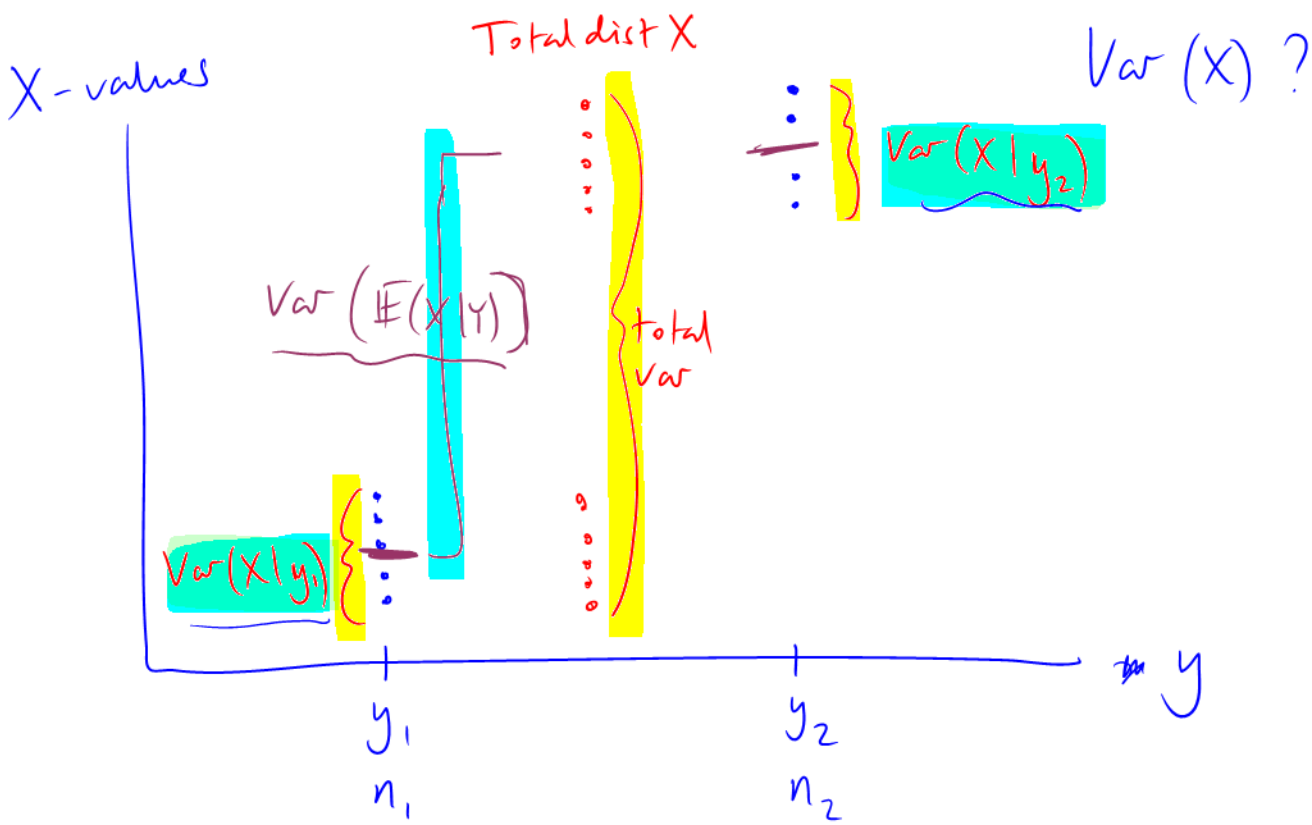
- The total average is $\mathbb{E}(X)$;
- The case-by-case averages are $\mathbb{E}(X \mid Y)$ for the different values of Y;
- The average of case-by-case averages is the average over Y of the Y-case averages: $\mathbb{E}_Y\{\mathbb{E}(X \mid Y)\}$.

"the mean of the means is equal to the mean!"



Total Expectation $= \underset{Y}{\text{mean}}$ (case-by-case expections)

$\mathbb{E}_Y$ not $\mathbb{E}Y$!!!



Example: In the example above, we had: $\mathbb{E}(X \mid Y) = \begin{cases} 9/4 & \text{with probability } 1/8, \\ 18/4 & \text{with probability } 7/8. \end{cases}$

The total average is:

$$\mathbb{E}(X) = \mathbb{E}_Y\{\mathbb{E}(X \mid Y)\} = \frac{9}{4} * \frac{1}{8} + \frac{18}{4} * \frac{7}{8} = 4.22.$$

Proof of (i), (ii), (iii): Prove $\mathbb{E}(g(X)) = \mathbb{E}_Y\{\mathbb{E}(g(X) \mid Y)\}$

(i) is a special case of (ii), so we just need to prove (ii). Begin at RHS:

$$\begin{aligned}
\text{RHS} &= \mathbb{E}_Y\left[\mathbb{E}(g(X) \mid Y)\right] = \mathbb{E}_Y\left[\sum_x g(x)\mathbb{P}(X = x \mid Y)\right] \\
&= \sum_y \left[\sum_x g(x)\mathbb{P}(X = x \mid Y = y)\right] \mathbb{P}(Y = y) \\
&= \sum_y \sum_x g(x)\mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y) \\
&= \sum_x g(x) \sum_y \mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y) \\
&= \sum_x g(x)\mathbb{P}(X = x) \qquad \text{(partition rule)} \\
&= \mathbb{E}(g(X)) \quad = \quad \text{LHS.}
\end{aligned}$$

$$\mathbb{E}(h(Y)) = \sum_y h(y)\,\mathbb{P}(Y=y)$$

(iii) Wish to prove $\text{Var}(X) = \mathbb{E}_Y[\text{Var}(X \mid Y)] + \text{Var}_Y[\mathbb{E}(X \mid Y)]$. Begin at RHS:

$\mathbb{E}_Y[\text{Var}(X \mid Y)] + \text{Var}_Y[\mathbb{E}(X \mid Y)]$

$$= \mathbb{E}_Y\left\{\mathbb{E}(X^2 \mid Y) - (\mathbb{E}(X \mid Y))^2\right\} + \left\{\mathbb{E}_Y\left\{[\mathbb{E}(X \mid Y)]^2\right\} - \Big[\underbrace{\mathbb{E}_Y(\mathbb{E}(X \mid Y))}_{\mathbb{E}(X) \text{ by part (i)}}\Big]^2\right\}$$

$$= \underbrace{\mathbb{E}_Y\{\mathbb{E}(X^2 \mid Y)\}}_{\mathbb{E}(X^2) \text{ by part (i)}} - \mathbb{E}_Y\{[\mathbb{E}(X \mid Y)]^2\} + \mathbb{E}_Y\{[\mathbb{E}(X \mid Y)]^2\} - (\mathbb{E}X)^2$$

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

$$= \text{Var}(X) = \text{LHS.} \qquad \square$$

no trip | Pois(μ) Pois(μ) Pois(μ) $Y=3$ | no trip | Pois(μ) $Y=1$ | no trip | P(μ) P(μ) P(μ) $Y=3$ | no trip

3.4 Examples of Conditional Expectation and Variance

$X = 63$ $\quad X = 18$ $\quad X = 58$ — e.g. $\mu = 20/\text{day}$

1. Swimming with dolphins

Fraser runs a dolphin-watch business. Every day, he is unable to run the trip due to bad weather with probability p, independently of all other days. Fraser works every day except the bad-weather days, which he takes as holiday.

Let Y be the number of consecutive days Fraser has to work between bad-weather days. Let X be the total number of customers who go on Fraser's trip in this period of Y days. Conditional on Y, the distribution of X is

$$(X \mid Y) \sim \text{Poisson}(\mu Y).$$

(a) Name the distribution of Y, and state $\mathbb{E}(Y)$ and $\text{Var}(Y)$.

(b) Find the expectation and the variance of the number of customers Fraser sees between bad-weather days, $\mathbb{E}(X)$ and $\text{Var}(X)$.

(a) Let "success" = "holiday" = "bad-weather day" : prob p
"failure" = "work day" = "good weather day"

Y = # failures before the ~~next~~ first success

So $Y \sim \text{Geometric}(p)$

$$\mathbb{E}(Y) = \frac{1-p}{p} \qquad \text{Var}(Y) = \frac{1-p}{p^2}.$$

e.g. $p = 1/4$, $\mathbb{E}(Y) = \frac{3/4}{1/4} = 3$

(b) We know $(X \mid Y) \sim \text{Poisson}(\mu Y)$

So $\mathbb{E}(X \mid Y) = \mu Y$ $\quad *$

and $\text{Var}(X \mid Y) = \mu Y$

By the Law of Total Expectation:

$$\mathbb{E}(X) = \mathbb{E}_Y\{\mathbb{E}(X \mid Y)\}$$
$$= \mathbb{E}_Y\{\mu Y\}$$
$$= \mathbb{E}(\mu Y)$$
$$= \mu \mathbb{E} Y$$
$$\therefore \mathbb{E}(X) = \mu \frac{(1-p)}{p}.$$

can drop $\mathbb{E}_Y$ now because Y is the only thing it could be!

By the Law of Total Variance:

$$\mathrm{Var}(X) = \mathbb{E}_Y\{\mathrm{Var}(X \mid Y)\} + \mathrm{Var}_Y\{\mathbb{E}(X \mid Y)\}$$
$$= \mathbb{E}_Y(\mu Y) + \mathrm{Var}_Y(\mu Y)$$
$$= \mu \mathbb{E}(Y) + \mu^2 \mathrm{Var}(Y)$$
$$= \mu\left(\frac{1-p}{p}\right) + \mu^2\left(\frac{1-p}{p^2}\right) \quad \text{by (a)}$$
$$= \underbrace{\left(\frac{1-p}{p^2}\right)}_{\mathrm{Var}(Y)} \mu(p+\mu)$$

Checking your answer in R:

If you know how to use a statistical package like R, you can check your answer to the question above as follows.

```
> # Pick a value for p, e.g. p = 0.2.
> # Pick a value for mu, e.g. mu = 25
>
> # Generate 10,000 random values of Y ~ Geometric(p = 0.2):
> y <- rgeom(10000, prob=0.2)
>
> # Generate 10,000 random values of X conditional on Y:
> # use (X | Y) ~ Poisson(mu * Y) ~ Poisson(25 * Y)
> x <- rpois(10000, lambda = 25*y)
```

$y = (3, 3, 0, 1, 3, 2, \ldots)$

$x = (63, 18, 58, \ldots)$

mean(x) var(x)

```
> # Find the sample mean of X (should be close to E(X)):
> mean(x)
[1] 100.6606
>
> # Find the sample variance of X (should be close to var(X)):
> var(x)
[1] 12624.47
>
> # Check the formula for E(X):
> 25 * (1 - 0.2) / 0.2
[1] 100
>
> # Check the formula for var(X):
> 25 * (1 - 0.2) * (0.2 + 25) / 0.2^2
[1] 12600
```

H_0 : mean = 100

The formulas we obtained by working give $\mathbb{E}(X) = 100$ and $\text{Var}(X) = 12600$. The sample mean was $\overline{x} = 100.6606$ (close to 100), and the sample variance was 12624.47 (close to 12600). Thus our working seems to have been correct.

---

2. Randomly stopped sum

e.g. N = customers

This model arises very commonly in stochastic processes. A random number N of events occur, and each event i has associated with it some cost, penalty, or reward X_i. The question is to find the mean and variance of the total cost / reward: $T_N = X_1 + X_2 + \ldots + X_N$.

e.g. amount money withdrawn

The difficulty is that the number N of terms in the sum is itself random.

X_i = money for customer i

T_N is called a randomly stopped sum: it is a sum of X_i's, randomly stopped at the random number of N terms. N=#custs

Example: Think of a cash machine, which has to be loaded with enough money to cover the day's business. The number of customers per day is a random number N. Customer i withdraws a random amount X_i. The total amount withdrawn during the day is a randomly stopped sum: $T_N = X_1 + \ldots + X_N$.

sum is randomly stopped at N terms

Cash machine example

The citizens of Remuera withdraw money from a cash machine according to the following probability function (X):

Amount, x (\$)	50	100	200
$\mathbb{P}(X = x)$	0.3	0.5	0.2

The number of customers per day has the distribution $N \sim \text{Poisson}(\lambda)$.

Let $T_N = X_1 + X_2 + \ldots + X_N$ be the total amount of money withdrawn in a day, where each X_i has the probability function above, and $X_1, X_2, \ldots$ are independent of each other and of N.

T_N is a randomly stopped sum, stopped by the random number of N customers.

(a) Show that $\mathbb{E}(X) = 105$, and $\text{Var}(X) = 2725$.

(b) Find $\mathbb{E}(T_N)$ and $\text{Var}(T_N)$: the mean and variance of the amount of money withdrawn each day.

Solution

(a) Exercise.

(b) Let $T_N = \sum_{i=1}^{N} X_i$. If we knew N, it would be easy to find $\mathbb{E}(T_N)$ and $\text{Var}(T_N)$ as mean & var of a sum of indept r.v.s

So "pretend" we know N, and use conditional $\mathbb{E}(\cdot \mid N)$ and $\text{Var}(\cdot \mid N)$.

i.e. CONDITION on N.

$$\mathbb{E}(T_N \mid N) = \mathbb{E}(X_1 + X_2 + \cdots + X_N \mid N)$$

$= \mathbb{E}(X_1 + \cdots + X_N)$ because all X_i's are indept of N, where N is now treated as a constant

$= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_N)$ don't need indep for this

$= N * \mathbb{E}(X)$ because all X_i's have the same mean, $\mathbb{E}X = 105$

$$\mathbb{E}(T_N \mid N) = 105N.$$

$N \sim \text{Poisson}(\lambda)$

$\therefore \mathbb{E}(T_N) = \mathbb{E}_N\{\mathbb{E}(T_N \mid N)\} = \mathbb{E}_N(105N) = 105\,\mathbb{E}N = 105\lambda$

Similarly,

$Var(T_N \mid N) = Var(X_1 + \dots + X_N \mid N)$

$= Var(X_1 + \dots + X_N)$ where N is now considered constant (because X_i's are independent of N).

$= Var(X_1) + \dots + Var(X_N)$ because X_i's are indept of each other. 1mark

$= N\,Var(X)$ because all X_i's have the same variance which is $Var(X)$.

$\therefore$ $Var(T_N \mid N) = 2725N.$ (using (a)).

$\therefore$ $Var(T_N) = \mathbb{E}_N\{Var(T_N \mid N)\} + Var_N\{\mathbb{E}(T_N \mid N)\}$

$= \mathbb{E}_N\{2725N\} + Var_N\{105N\}$

$= 2725\,\mathbb{E}(N) + 105^2\,Var(N)$

$N \sim \text{Poisson}(\lambda)$
so $\mathbb{E}(N) = Var(N) = \lambda$

$= 2725\lambda + 105^2\lambda$

$= 13750\lambda$

For a quick ~~(crude)~~ way to cover $\simeq 97.5\%$ of days' business, use

Loaded # Money $= \mathbb{E}(T_N) + 1.96\sqrt{Var(T_N)}$

$= 105\lambda + 1.96\sqrt{13750\lambda}$

Central Limit Theorem.

Check in R (advanced)

```
> # Create a function tn.func to calculate a single value of T_N
> # for a given value N=n:
> tn.func <- function(n){
        sum(sample(c(50, 100, 200), n, replace=T,
        prob=c(0.3, 0.5, 0.2)))
  }

> # Generate 10,000 random values of N, using lambda=50:
> N <- rpois(10000, lambda=50)
> # Generate 10,000 random values of T_N, conditional on N:
> TN <- sapply(N, tn.func)
> # Find the sample mean of T_N values, which should be close to
> # 105 * 50 = 5250:
> mean(TN)
[1] 5253.255
> # Find the sample variance of T_N values, which should be close
> # to 13750 * 50 = 687500:
> var(TN)
[1] 682469.4
```

All seems well. Note that the sample variance is often some distance from the true variance, even when the sample size is 10,000.

General result for randomly stopped sums:

Suppose $X_1, X_2, \ldots$ each have the same mean μ and variance σ^2, and $X_1, X_2, \ldots,$ and N are mutually independent. Let $T_N = X_1 + \ldots + X_N$ be the randomly stopped sum. By following similar working to that above:

$$\underbrace{\mathbb{E}(T_N)}_{\text{mean total}} = \mathbb{E}\left\{\sum_{i=1}^{N} X_i\right\} = \underbrace{\mu}_{\text{mean per cust}}\,\underbrace{\mathbb{E}(N)}_{\text{Mean \# customers}}$$

Wald's Equation. 721 Ass 2

$$\mathrm{Var}(T_N) = \mathrm{Var}\left\{\sum_{i=1}^{N} X_i\right\} = \underbrace{\sigma^2\,\mathbb{E}(N)}_{\text{cpt due to var of X's "within SS"}} + \underbrace{\mu^2\,\mathrm{Var}(N)}_{\text{cpt due to var in N "between SS"}}.$$

$\mathbb{E}(\text{time to end } A) = \infty$

$\mathbb{E}(\text{time to either end}) = \text{interesting}$

$$\boxed{\text{End } A} \leftarrow \boxed{\text{Process}} \rightarrow \boxed{\text{End } B}$$

??

3.5 First-Step Analysis for calculating expected reaching times

Remember from Section 2.6 that we use First-Step Analysis for finding the probability of eventually reaching a particular state in a stochastic process. First-step analysis for probabilities uses conditional probability and the Partition Theorem (Law of Total Probability).

In the same way, we can use first-step analysis for finding the expected reaching time for a state.

$$\boxed{\text{Process}} \longrightarrow \boxed{\text{Stop}} \qquad \mathbb{E}(\text{time to stop})?$$

This is the expected number of steps that will be needed to reach a particular state from a specified start-point, or the expected length of time it will take to get there if we have a continuous time process.

Just as first-step analysis for probabilities uses conditional probability and the law of total probability (Partition Theorem), first-step analysis for expectations uses conditional expectation and the Law of Total Expectation.

First-step analysis for probabilities:

The first-step analysis procedure for probabilities can be summarized as follows:

$$\mathbb{P}(\text{eventual goal}) = \sum_{\substack{\text{1st-step}\\ \text{options}}} \mathbb{P}(\text{eventual goal} \mid \text{option})\, \mathbb{P}(\text{option})$$

Sample space $= \Omega = \{$all routes from current position to end$\}$

This is because the first-step options form a partition of the sample space Ω.

First-step analysis for expected reaching times:

The expression for expected reaching times is very similar:

$$\mathbb{E}(\text{reaching time}) = \sum_{\substack{\text{1st step}\\ \text{options}}} \mathbb{E}(\text{reaching time} \mid \text{option})\, \mathbb{P}(\text{option})$$

Tomorrow's Office Hour: 3.30 – 4.

$$\mathbb{E}_Y(h(Y)) = \sum_y h(y)\,\mathbb{P}(Y=y)$$

This follows immediately from the law of total expectation:

$$\mathbb{E}(X) = \mathbb{E}_Y\{\mathbb{E}(X \mid Y)\} = \sum_y \mathbb{E}(X \mid Y=y)\,\mathbb{P}(Y=y)$$

Let X be the reaching time, and let Y be the label for possible 1st-step options:

i.e. $Y = 1, 2, 3 \ldots$ for options $1, 2, 3, \ldots$

We then obtain:

$$\mathbb{E}(X) = \sum_y \mathbb{E}(X \mid Y=y)\,\mathbb{P}(Y=y)$$

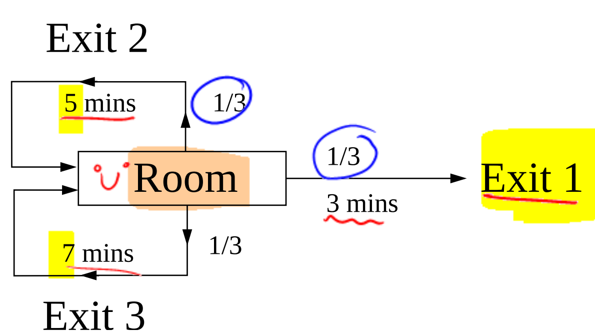
i.e. $\mathbb{E}(\text{reaching time}) = \sum_{\substack{\text{1st-step} \\ \text{options}}} \mathbb{E}(\text{reaching time} \mid \text{option})\,\mathbb{P}(\text{option})$.

Example: Mouse in a maze

A mouse is trapped in a room with three exits at the centre of a maze.

- Exit 1 leads outside the maze after 3 minutes.
- Exit 2 leads back to the room after 5 minutes.
- Exit 3 leads back to the room after 7 minutes.

Every time the mouse makes a choice, it is equally likely to choose any of the three exits. What is the expected time taken for the mouse to leave the maze?



Define notation: let

$$m_R = \mathbb{E}\left(\begin{matrix}\text{time} \\ \text{to end}\end{matrix} \;\middle|\; \begin{matrix}\text{start in} \\ \text{Room}\end{matrix}\right)$$

First-step analysis:

$$m_R = \frac{1}{3}\underbrace{\mathbb{E}(\text{time} \mid \text{Exit } 1)}_{3\text{ mins}} + \frac{1}{3}\underbrace{\mathbb{E}(\text{time} \mid \text{Exit } 2)}_{5\text{ mins} + m_R} + \frac{1}{3}\underbrace{\mathbb{E}(\text{time} \mid \text{Exit } 3)}_{7 + m_R}$$

$$\therefore m_R = \frac{1}{3} * 3 + \frac{1}{3}(5 + m_R) + \frac{1}{3}(7 + m_R)$$

$$\therefore m_R\left(1 - \frac{1}{3} - \frac{1}{3}\right) = 1 + \frac{5}{3} + \frac{7}{3}$$

$$\therefore \quad \underline{m_R = 15 \text{ mins}}$$

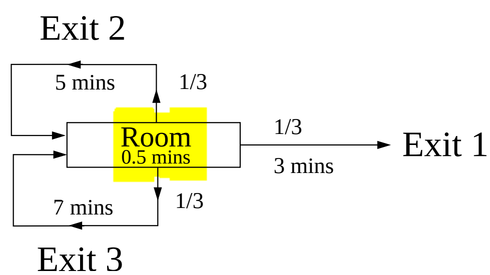
Expected time to get out, starting from the room.

(Heavy notation way:
$\mathbb{E}(X \mid Y=1) = 3$ mins
$\mathbb{E}(X \mid Y=2) = 5 + \mathbb{E}(X)$ etc.)

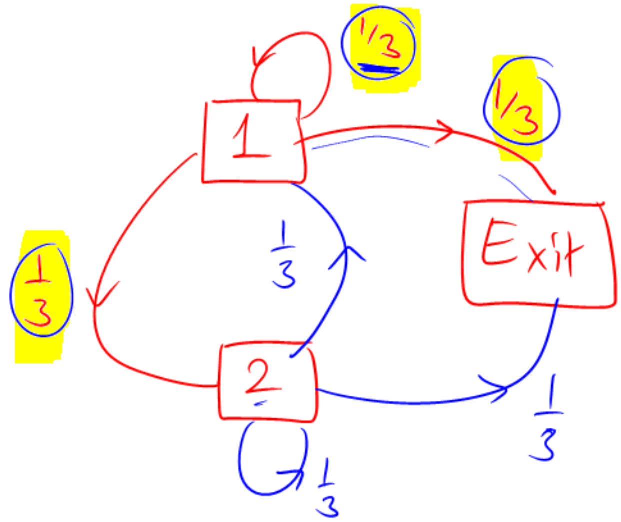
Don't use

Two approaches to first-step analysis for expectations

In some problems, all possible first-step options incur the same initial time penalty. For example, in the mouse example, suppose that the mouse always takes 0.5 mins to choose an exit every time it enters the room.



Another Example:



~~3~~ 3 steps.

Find $\mathbb{E}$(# steps to finish | start at Room 1)
mean

Exp. reaching time for state "Exit"

Counting arrows traversed until the end.

Define $\to m_1 = \mathbb{E}$(# steps to finish | start at state 1) ②
$m_2 = \mathbb{E}$(# ———— | ———— 2)

FSA: $m_1 = 1 + \frac{1}{3} m_1 + \frac{1}{3} m_2 + \frac{1}{3} * 0$ ← * easier ①

Same thing OR $m_1 = \frac{1}{3}(1 + m_1) + \frac{1}{3}(1 + m_2) + \frac{1}{3}(1 + 0)$

if different directions "cost" different amounts, need this one ②

And: $m_2 = 1 + \frac{1}{3} m_1 + \frac{1}{3} m_2 + \frac{1}{3} * 0$ ②

Solve ① and ② $\Rightarrow$ $m_1 = 3$ steps. $= m_2$. (CHECK)

Calculator Average # steps to get out starting from box 1 ①

In this case, we have two choices:

1. Add the extra 0.5 mins onto each option separately:

$$m_R = \frac{1}{3}(0.5 + 3) + \frac{1}{3}(0.5 + 5 + m_R) + \frac{1}{3}(0.5 + 7 + m_R)$$

Better: 2. Add the 0.5 mins on at the begining:

$$m_R = 0.5 + \frac{1}{3}(3) + \frac{1}{3}(5 + m_R) + \frac{1}{3}(7 + m_R).$$

In each case, we will get the same answer (check). This is because the option probabilities sum to 1.

3.6 Probability as a conditional expectation

Recall from Section 3.1 that for any event A, we can write $\mathbb{P}(A)$ as an expectation as follows.

Define the indicator random variable: $I_A = \begin{cases} 1 & \text{if event } A \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$ Prob: $\mathbb{P}(A)$, $1 - \mathbb{P}(A)$

Then $\mathbb{E}(I_A) = \mathbb{P}(I_A = 1) = \mathbb{P}(A)$.

$$\mathbb{E}(I_A) = 1 * \mathbb{P}(A) + 0 * (1 - \mathbb{P}(A))$$

We can refine this expression further, using the idea of conditional expectation. Let Y be any random variable. Then

$$\mathbb{P}(A) = \mathbb{E}(I_A) = \mathbb{E}_Y\{\mathbb{E}(I_A \mid Y)\}$$

Law of Total Expectation §3.3.

But

$$\mathbb{E}(I_A \mid Y) = \sum_{r=0}^{1} r\,\mathbb{P}(I_A = r \mid Y)$$

$$= 0 * \mathbb{P}(I_A = 0 \mid Y) + 1 * \mathbb{P}(I_A = 1 \mid Y)$$

$$= \mathbb{P}(I_A = 1 \mid Y)$$

$$= \mathbb{P}(A \mid Y).$$

Probability as a conditional expectation

Therefore: $\mathbb{P}(A) = \mathbb{E}_Y\{\mathbb{E}(I_A \mid Y)\} = \mathbb{E}_Y\{\mathbb{P}(A \mid Y)\}$.

Thus if Y is discrete: $\mathbb{P}(A) = \sum_y \mathbb{P}(A \mid Y=y)\,\mathbb{P}(Y=y)$ $\boxed{\text{Partition Thm}}$

if Y is continuous: $\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid Y=y)\, f_Y(y)\, dy$ $\boxed{\text{New}}$

This means that for ***any*** random variable X (discrete or continuous), and for any set of values S (a discrete set or a continuous set), we can write:

event $A = \{X \in S\}$

- for any ***discrete*** random variable Y,

$$\mathbb{P}(X \in S) = \sum_y \mathbb{P}(X \in S \mid Y=y)\,\mathbb{P}(Y=y)$$

- for any ***continuous*** random variable Y,

$$\mathbb{P}(X \in S) = \int_{-\infty}^{\infty} \mathbb{P}(X \in S \mid Y=y)\, f_Y(y)\, dy$$

where $f_Y(y)$ is PDF of Y.

Example of probability as a conditional expectation: winning a lottery

Suppose that a million people have bought tickets for the weekly lottery draw. Each person has a probability of one-in-a-million of selecting the winning numbers. If more than one person selects the winning numbers, the winner will be chosen at random from all those with matching numbers.

You watch the lottery draw on TV and your numbers match the winning numbers!!! Only a one-in-a-million chance, and there were only a million players, but what is your probability of winning the prize....?

Define X to be the number of matching tickets out of the 1 million tickets sold. (If you are lucky, $X = 1$ and it's you!)

If there are 1 million tickets and each ticket has a one in a million chance of having the winning numbers, then

The relationship $X \sim \text{Poisson}(1)$ arises because of the Poisson approximation to the Binomial distribution.

(a) You have a matching ticket, so we know that . Let Y be the number of matching tickets, given that there is at least one. Find the probability function of Y, $f_Y(y)$.

(b) What is the probability that there is only one matching ticket, given that there is at least one (i.e. yours)?

(c) The prize is chosen at random from all those who have matching tickets. What is the probability that you win?

Using a computer, this probability evaluates to
Disappointing. . . ?

Whole example, sample space $\Omega = \{$all draws where I have matching ticket$\}$

Example of probability as a conditional expectation: winning a lottery

Suppose that a million people have bought tickets for the weekly lottery draw. Each person has a probability of one-in-a-million of selecting the winning numbers. If more than one person selects the winning numbers, the winner will be chosen at random from all those with matching numbers.

You watch the lottery draw on TV and your numbers match the winning numbers!!! Only a one-in-a-million chance, and there were only a million players, but what is your probability of winning the prize....?

Define Y to be the number of OTHER matching tickets out of the OTHER 1 million tickets sold. (If you are lucky, $Y = 0$ so you are the only person who can win the prize!)

If there are 1 million tickets and each ticket has a one in a million chance of having the winning numbers, then

$$Y \sim \text{Poisson}\left(\frac{1 \text{ mill}}{1 \text{ mill}}\right) = \text{Poisson}(1) \text{ approximately.}$$

The relationship $Y \sim \text{Poisson}(1)$ arises because of the Poisson approximation to the Binomial distribution. [Exactly, $Y \sim \text{Bin}\left(1 \text{ mill}, \frac{1}{1 \text{ mill}}\right)$.]

(a) What is the probability function of Y, $f_Y(y)$? $Y \sim \text{Poisson}(1)$.

$$f_Y(y) = \mathbb{P}(Y = y) = \frac{1^y e^{-1}}{y!} = \frac{1}{e * y!} \quad \text{for } y = 0, 1, 2, \ldots$$

(b) What is the probability that yours is the only matching ticket?

$$\mathbb{P}(Y = 0) = \frac{1}{e * 0!} = \frac{1}{e} = \boxed{0.368}.$$

$e = e^1 = e^y$ 1 on calculator

(c) The prize is chosen at random from all those who have matching tickets. What is the probability that you win if there are $Y = y$ OTHER matching tickets?

Let W be the event $W = \{\text{I win}\}$.

$$\mathbb{P}(W \mid Y = y) = \frac{1}{y+1}.$$

$y = 0$	1
$y = 1$	$1/2$
$y = 2$	$1/3$

$\Omega = \{ \text{I match} \}$

(d) Overall, what is the probability that you win, given that you have a matching ticket?

$$\mathbb{P}_{match}(W) = \mathbb{E}_Y \{ \mathbb{P}(W \mid Y=y) \} \text{ from p.65.}$$

$$= \sum_{y=0}^{\infty} \mathbb{P}(W \mid Y=y)\, \mathbb{P}(Y=y)$$

$$= \sum_{y=0}^{\infty} \left(\frac{1}{y+1}\right)\left(\frac{1}{e * y!}\right)$$

$$= \frac{1}{e} \sum_{y=0}^{\infty} \frac{1}{(y+1)\, y!} \qquad (y+1)!$$

$$= \frac{1}{e} \sum_{y=0}^{\infty} \frac{1}{(y+1)!} = \frac{1}{e}\left\{ \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots \right\}$$

[Let $x = y+1$]

$$= \frac{1}{e}\left\{ \sum_{x=0}^{\infty} \frac{1}{x!} - \frac{1}{0!} \right\} = \frac{1}{e}\left\{ \underbrace{\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \dots}\ - \frac{1}{0!} \right\}$$

$$= \frac{1}{e}\{ e - 1 \}$$

$$= 1 - \frac{1}{e}$$

$$= 0.632.$$

Poisson(1) probs: $\sum_{x=0}^{\infty} \frac{1^x}{x!} e^{-1} = 1$

$\therefore \sum_{x=0}^{\infty} \frac{1}{x!} = e$

Disappointing. . . ?