Mixture of Discrete and Continuous Random Variables

- What does the CDF $F_X(x)$ look like when X is discrete vs when it's continuous?
- A r.v. could have a continuous component and a discrete component.
- Ex 1 & 2 from `MixedRandomVariables.pdf`.

Example 1: Consider a r.v. X with cdf

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{3} & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

The support of X is $[0, 2] = A_1 \cup A_2$ where $A_1 = [0, 2)$[a], and $A_2 = \{2\}$. The distribution of X has different expressions over the two regions:

- (continuous portion) pdf on A_1 with $f(x) = 1/3$.
- (discrete portion) pmf on A_2, with $p(2) = 1/3$.

When computing expectations, we use pmf or pdf, in each region.

[a] It doesn't matter if we write $A_1 = (0, 2)$.

Example 2

Suppose $a > b$. The difference between the two sets, $(-\infty, b]$ and $(-\infty, a]$, is $(b, a]$. So

$$
\begin{aligned}
F(a) - F(b) &= \mathbb{P}(X \leq a) - \mathbb{P}(X \leq b) \\
&= \mathbb{P}\big(X \in (-\infty, a]\big) - \mathbb{P}\big(X \in (-\infty, b]\big) \\
&= \mathbb{P}\big(X \in (b, a]\big) \\
&= \mathbb{P}(b < X \leq a).
\end{aligned}
$$

How to use CDF to compute those probabilities? Note

1) $\mathbb{P}(b < X \leq a) = F(a) - F(b)$;

2) if a point x_0 is in the continuous portion, then we can use $\geq$ and $>$ (or, $\leq$ and $<$) interchangeably.

- $\mathbb{P}(1 < X < 1.5) = \mathbb{P}(1 < X \leq 1.5) = F(1.5) - F(1) - 0$, where the first equality is due to the fact that 1.5 is in the continuous portion.
- $\mathbb{P}(1 \leq X < 1.5) = \mathbb{P}(1 \leq X \leq 1.5) = \mathbb{P}(X = 1) + \mathbb{P}(1 < X \leq 1.5) = 0.5 + F(1.5) - F(1)$.
- $\mathbb{P}(1 \leq X \leq 1.5) = \mathbb{P}(X = 1) + \mathbb{P}(1 < X \leq 1.5)$.

Example: An insurance policy reimburses a loss up to a benefit limit of C and has a deductible of d. Suppose the policyholder's loss, X, follows Ex$(1/5)$. Let Y denote the benefit paid under the insurance policy. Find the distribution of Y.

$$Y = \text{Benefit Paid} = \begin{cases} 0 & x < d \\ x - d & d \le x < C + d \\ C & x \ge C + d \end{cases}$$

- Discrete portion of Y:

$$p(0) = \int_0^d \frac{1}{5} e^{-x/5} dx = 1 - e^{-d/5}, \quad p(C) = \int_{C+d}^{\infty} \frac{1}{5} e^{-x/5} dx = e^{-(C+d)/5}.$$

- Continuous portion of Y:

$$f_Y(y) = \frac{1}{5} e^{-(y+d)/5}, \quad 0 < y < C.$$

Distributions of Two Random Variables

Major concepts (chap 2):

- Joint pdf/pmf
- Marginal pdf/pmf
- Conditional pdf/pmf, conditional expectations

- Let X and Y be discrete random variables. The **joint pmf** $p(x, y)$ is defined by

$$p(x, y) = \mathbb{P}(X = x, Y = y),$$

and

$$\mathbb{P}((X, Y) \in A) = \sum\sum_{(x,y)\in A} p(x, y).$$

- The **marginal pmfs** of X and of Y are given by

$$p_X(x) = \sum_{\text{all } y} p(x, y), \quad p_Y(y) = \sum_{\text{all } x} p(x, y).$$

Check that $p_X(x)$ and $p_Y(y)$ are legit pmfs.

Example 1 (2.1.1 on p.74)

X / Y	0	1	2	3
0	$\frac{1}{8}$	$\frac{1}{8}$	0	0
1	0	$\frac{2}{8}$	$\frac{2}{8}$	0
2	0	0	$\frac{1}{8}$	$\frac{1}{8}$

a) Find $\mathbb{P}(X+Y=2)$. $p(1,1)+p(0,2)+p(2,0)=\frac{2}{8}$.

b) Find $\mathbb{P}(X<Y)$.

c) Find the marginal probability distributions $p_X(x)$ of X and $p_Y(y)$ of Y. For example, the marginal pmf of X is given by the row sums of the table

$$p_X(0)=\frac{1}{8}+\frac{1}{8}=\frac{1}{4},\quad p_X(1)=\frac{2}{8}+\frac{2}{8}=\frac{1}{4},\quad p_X(2)=\frac{1}{4}.$$

- Let X and Y be continuous random variables. Then $f(x, y)$ is the **joint pdf** for X and Y if for any two-dimensional set A

$$\mathbb{P}((X, Y) \in A) = \iint_A f(x, y) dx dy.$$

In particular, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

- The **marginal pdfs** of X and of Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Check that $f_X(x)$ is a legit pdf: apparently $f_X(x) \geq 0$. and

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1.$$

- Let X and Y be random variables. Then their **joint cdf** is defined by

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

For continuous rvs,

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

- Let X and Y be random variables and g some real valued function, i.e., $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$.

$$\mathbb{E}[g(X, Y)] = \sum_{\text{all } x} \sum_{\text{all } y} g(x, y) \cdot p(x, y)$$

$$\mathbb{E}[g(X, Y)] = \iint_{\mathbb{R}^2} g(x, y) f(x, y) dx dy.$$

- The **mgf** of a random vector (X, Y)

$M_{XY}(t_1, t_2) = \mathbb{E}(e^{t_1 X + t_2 Y})$, if it exists for $|t_1| < h, |t_2| < h$.

$$\begin{aligned} M_X(t) &= M_{XY}(t, 0) \\ M_Y(t) &= M_{XY}(0, t). \end{aligned}$$

Example 2: Consider two random variables X and Y with the mgf

$$M(t_1, t_2) = 0.10 + 0.20e^{t1} + 0.30e^{2t_2} + 0.40e^{t_1+t_2}. \tag{1}$$

Find the joint pmf $p(x, y)$.

Recall that $M(t_1, t_2) = \mathbb{E}(e^{t_1 X + t_2 Y}) = \sum_{\text{all } (x,y)} p(x, y)e^{x \cdot t_1 + y \cdot t_2}$.

Write (1) as

$$0.10e^{(0)t_1+(0)t_2} + 0.20e^{(1)t1+(0)t_2} + 0.30e^{(0)t_1+2t_2} + 0.40e^{(1)t_1+(1)t_2}.$$

So

X / Y	0	1	2
0	0.10	0	0.30
1	0.20	0.40	0

Example 3 (2.1.5 & 2.1.6 on p.80-81) Let X and Y have the pdf

$$f(x, y) = 8xy, \quad 0 < x < y < 1; \quad 0, \text{ elsewhere.}$$

The very first step before doing any calculation: sketch the support.

a) Verify that $f(x, y)$ is a legitimate pdf.

1. $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$.

2. $\int\int f(x, y)dxdy = 1$ — this is because

$$\begin{aligned} \int_0^1 \int_0^y 8xydxdy &= \int_0^1 8y\Big(\int_0^y xdx\Big)dy \\ &= \int_0^1 4y^3 dy \quad \text{where } \int_0^y xdx = \frac{x^2}{2}\Big|_0^y = \frac{y^2}{2} \\ &= y^4\Big|_0^1 = 1. \end{aligned}$$

b) Find $\mathbb{P}(X+Y<0.5)$ and $\mathbb{P}(2X \geq Y)$.

$$\begin{aligned}
\mathbb{P}(X+Y<0.5) &= \int_0^{0.25}\Big(\int_x^{0.5-x} 8xydy\Big)dx \\
&= \int_0^{0.25} 4x\Big(\int_x^{0.5-x} 2ydy\Big)dx \\
&= \int_0^{0.25} 4x\Big[(0.5-x)^2 - x^2\Big]dx \\
&= \int_0^{0.25} 4x(0.25-x)dx = \int_0^{0.25}(x-4x^2)dx \\
&= \frac{0.25^2}{2} - \frac{4}{3}0.25^3
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(2X \geq Y) &= \int_0^{0.5}\Big(\int_x^{2x} 8xy\ dy\Big)dx + \int_{0.5}^{1}\Big(\int_x^{1} 8xy\ dy\Big)dx, OR \\
&= \int_0^{1}\Big(\int_{y/2}^{y} 8xy\ dx\Big)dy
\end{aligned}$$

c) Find the marginal pdf for X.

$$f_X(x) = \int_0^1 f(x,y)dy = \int_x^1 8xy\ dy = 4x - 4x^2, \quad 0 < x < 1.$$

d) Find the marginal pdf for Y.

$$f_X(x) = \int_0^1 f(x,y)dx = \int_0^y 8xy\ dx = 4y^3, \quad 0 < y < 1.$$

e) Find $\mathbb{E}(XY^2)$, $\mathbb{E}(Y)$, $\mathbb{E}(7XY^2 + 5Y)$ (see p.80).

f) Let $Z = X/Y$. Find the distribution of Z.

First find the support of Z

$$0 < x < y < 1 \quad \Longrightarrow 0 < z < 1.$$

Then use the CDF approach

$$\begin{aligned} F_Z(z) &= \mathbb{P}(X/Y \leq z) = \mathbb{P}(X \leq zY) \\ &= \int_0^1 \Big(\int_0^{zy} 8xy \; dx \Big) dy = \int_0^1 4y^3 z^2 dy \\ &= z^2. \end{aligned}$$

The pdf is given by

$$f_Z(z) = 2z, \quad 0 < z < 1.$$

Go through the examples from `JointDistributions.pdf` by yourself.

Independent Random Variables

- Random variables X and Y are **independent** if for all (x, y)

$$\begin{aligned} p(x,y) &= p_X(x) \cdot p_Y(y), \qquad \text{(discrete)} \\ f(x,y) &= f_X(x) \cdot f_Y(y), \qquad \text{(continuous).} \end{aligned}$$

- If X and Y independent,
  1. $F(x, y) = F_X(x) \cdot F_Y(y)$
  2. $M(t_1, t_2) = M_X(t_1) \cdot M_Y(t_2)$
  3. $\mathbb{E}[g(X)h(Y)] = \mathbb{E}g(X) \cdot \mathbb{E}h(Y)$, specially $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ (see the proof on the next slide).

$$
\begin{aligned}
\mathbb{E}(X+Y) &= \mathbb{E}X + \mathbb{E}Y = \mu_X + \mu_Y \\
\mathsf{Var}(X+Y) &= \mathbb{E}\big(X+Y-(\mu_X+\mu_Y)\big)^2 = \mathbb{E}\big(X-\mu_X+Y-\mu_Y\big)^2 \\
&= \mathbb{E}\Big[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)\Big] \\
&= \mathbb{E}(X-\mu_X)^2 + \mathbb{E}(Y-\mu_Y)^2 + 2\mathbb{E}(X-\mu_X)(Y-\mu_Y) \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathbb{E}(X-\mu_X) \times \mathbb{E}(Y-\mu_Y) \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y) + 0
\end{aligned}
$$

Example 1 (revisit): Are X and Y independent? (NO)

Sign that they are dependent: some entries in pmf table are zero.

Example 3 (revisit): Are X and Y independent? (NO)

Sign that they are dependent: support is not rectangle.

More Examples from `Independence_and_Covariance.pdf`