The Beta distribution

A random variable Y has the *beta distribution* with parameters a and b if its pdf is of the form

$$f(y) = \frac{y^{a-1}(1-y)^{b-1}}{B(a,b)} \quad 0 < y < 1 \qquad (a,b>0).$$

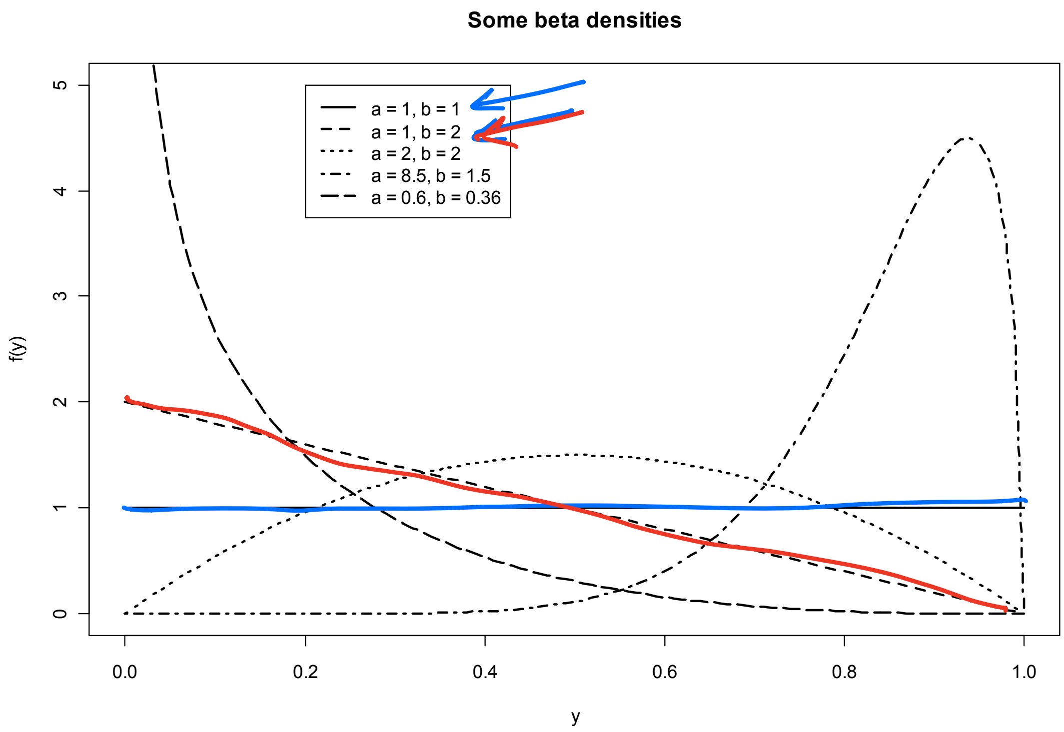
We write $Y \sim Beta(a,b)$ and $f(y) = f_{Beta(a,b)}(y)$.

Here, $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the *beta function*. Eg, $B(2,3) = \dfrac{\Gamma(2)\Gamma(3)}{\Gamma(2+3)} = \dfrac{1!2!}{4!} = \dfrac{1}{12}$.

A special case: If $a = b = 1$, then $f(y) = \dfrac{y^{1-1}(1-y)^{1-1}}{B(1,1)} = 1$, $0 < y < 1$.

Thus $Beta(1,1) = U(0,1)$.

It can easily be shown that $Mode(Y) = (a-1)/(a+b-2)$ if $a > 1$ and $b > 1$.



R Code (non-assessable)

```
yv = seq(0,1,0.01); X11(w=10,h=7)
plot(c(0,1),c(0,5),type="n",xlab="y",ylab="f(y)",main="Some beta
densities")
lines(yv,dbeta(yv,1,1),lty=1,lwd=2)
lines(yv,dbeta(yv,1,2),lty=2,lwd=2)
lines(yv,dbeta(yv,2,2),lty=3,lwd=2)
lines(yv,dbeta(yv,8.5,1.5),lty=4,lwd=2)
lines(yv,dbeta(yv,0.6,3.6),lty=5,lwd=2)
legend(0.2,5, c( "a = 1, b = 1","a = 1, b = 2","a = 2, b = 2",
       "a = 8.5, b = 1.5","a = 0.6, b = 0.36"),
       lty=c(1,2,3,4,5),lwd=c(2,2,2,2,2))
```

Expectation in the context of continuous distributions

Basically, all the definitions regarding expectation in Chapter 3 hold here also, except that *sums* need to be replaced by *integrals*.

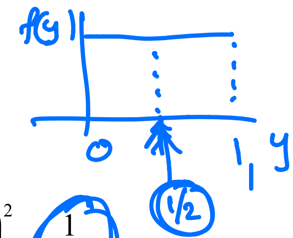
> If Y is a continuous random variable with pdf $f(y)$, and $g(t)$ is a function, then the expected value of $g(Y)$ is
> $Eg(Y) = \int g(y)f(y)dy$. (The integral is from minus infinity to infinity.)

As in Chapter 3:

$Ec = c,\ E\{cg(Y)\} = cEg(Y)$
$E\{g_1(Y) + \ldots + g_k(Y)\} = Eg_1(Y) + \ldots + Eg_k(Y)$ (3 laws of expectation)
$\mu = EY$ (mean = measure of central tendency)
$\mu_k' = EY^k = \int y^k f(y) dy$ (kth raw moment)
$\mu_k = E(Y-\mu)^k$ (kth central moment)
$\sigma^2 = \mu_2 = Var(Y)$ (variance = measure of dispersion)
$\sigma^2 = \mu_2' - \mu^2$ (formula for finding variances)
$Var(a+bY) = b^2 VarY$ (another such formula)
$m(t) = Ee^{Yt}$ (moment generating function)
$\mu_k' = m^{(k)}(0)$ (formula for finding moments)
$P(|Y-\mu| < k\sigma) \geq 1 - 1/k^2$ (Chebyshev's theorem)
$Mode(Y)$ = any value y such that the pdf $f(y)$ is a maximum
$Median(Y)$ = any value y such that $F(y) = 1/2$ (simpler than for discrete rvs).

Example 9 Find the mean and variance of the standard uniform distribution.

Suppose that $Y \sim U(0,1)$. Then Y has pdf $f(y) = 1,\ 0 < y < 1$.



So $\mu = \int_{-\infty}^{\infty} yf(y)dy = \int_0^1 y1dy = \left[\frac{y^2}{2}\Big|_{y=0}^{1}\right] = \frac{1}{2}$ $= \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}$

Also, $\mu_2' = \int_0^1 y^2 1dy = \left[\frac{y^3}{3}\Big|_{y=0}^{1}\right] = \frac{1}{3}$. Therefore $\sigma^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$.

(*Note*: We could also use the mgf method here, but it is problematic in this case. This is because, $m(t) = (e^t - 1)/t \Rightarrow m'(t) = \{e^t(t-1)+1\}/t^2$, which is *undefined* at $t = 0$.

So use *l'Hôpital's rule* (twice) to get $\mu = \lim_{t\to 0} m'(t) = \lim_{t\to 0}\left\{\frac{d\{e^t(t-1)+1\}/dt}{dt^2/dt}\right\}$

$= \lim_{t\to 0}\left\{\frac{te^t}{2t}\right\} = \lim_{t\to 0}\left\{\frac{d(te^t)/dt}{d(2t)/dt}\right\} = \lim_{t\to 0}\left\{\frac{e^t(t+1)}{2}\right\} = \frac{1}{2}$. This working is non-assessable.)

Example 10 Find the mean and variance of the exponential distribution.

$\frac{b}{1-bt} = c$

(In this case the mgf method works well.)

Suppose that $Y \sim Expo(b)$. Then Y has mgf

$$m(t) = \int_0^\infty e^{yt}\frac{1}{b}e^{-y/b}\,dy = \frac{1}{b}\int_0^\infty e^{-y\left(\frac{1}{b}-t\right)}dy$$

$= E e^{Yt}$

$$= \frac{1}{\cancel{b}}\left(\frac{\cancel{b}}{1-bt}\right)\int_0^\infty \left(\frac{1-bt}{b}\right)e^{-y\left(\frac{1-bt}{b}\right)}dy$$

pdf of an $Expo\left(\frac{b}{1-bt}\right)$

$\frac{1}{c}e^{-y/c}$

$(*)$

(where the integrand will be recognised as an exponential density, implying that the integral is 1)

$$m(t) = \boxed{(1-bt)^{-1}}.$$

So $m'(t) = -(1-bt)^{-2}(-b) = b(1-bt)^{-2}$.

Therefore $\mu = m'(0) = b(1-b0)^{-2} = b$. soln 1

Also, $m''(t) = -2b(1-bt)^{-3}(-b)$.

So $\mu_2' = m''(0) = 2b^2$.

So $\sigma^2 = 2b^2 - (b)^2 = b^2$.

IBP

Alternatively, we could use *integration by parts* to get the required moments directly:

$$\mu = \int_0^\infty y\left(\frac{1}{b}e^{-y/b}\right)dy = \ldots = b, \quad \mu_2' = \int_0^\infty y^2\left(\frac{1}{b}e^{-y/b}\right)dy = \ldots = 2b^2$$

exercise soln 2

soln 3: Use the trick at $(*)$ but not the mgf method & not IBP

What about Y's mode and median?

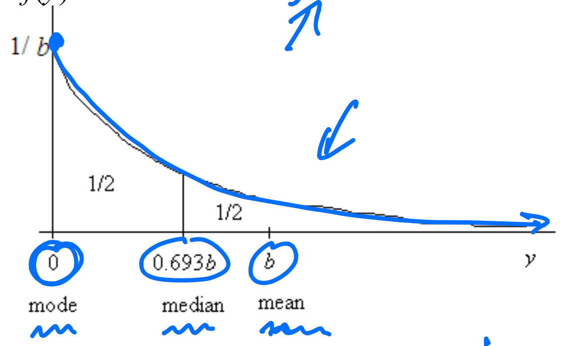
$Mode(Y) = 0$.

Y's median is the solution of $\boxed{F(y) = 1/2.}$

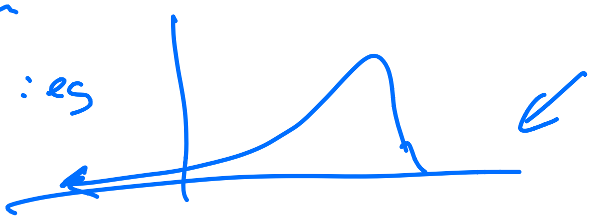
We set $1 - e^{-y/b} = 1/2$ and solve for y. ($F(y)$ was derived in Example 8.)

The result is $Median(Y) = b\log 2 = 0.693b$.

right skewed dsn



left skewed dsn : eg



Summary of continuous distributions

As an exercise, fill in the empty cells, and check against the back inside cover of text. You may wish to add two more columns, one for the mode and one for the median, although not all of these have a simple a formula.

distribution $Y\sim$	**pdf** $f(y)$	**mgf** $m(t)=Ee^{Yt}$	**mean** $\mu=EY$	**variance** $\sigma^2=VarY$
Uniform a b				
Standard uniform	$f(y)=1,\ 0<y<1$	$\boxed{}$	1/2	1/12
Normal				
Standard normal				
Gamma Gam(a,b)				
Chi-square $\chi^2(n)$ $=?$				
Exponential Expo(b) = Gam(1,b)	$\frac{1}{b}e^{-y/b},\ y>0$	$(1-bt)^{-1}$	b	b^2
Standard exponential				
Beta				

Completed summary of continuous distributions

distribution $Y \sim$	**pdf** $p(y)$	**mgf** $m(t) = Ee^{Yt}$	**mean** $\mu = EY$	**variance** $\sigma^2 = VarY$
Uniform $= Beta(1,1)$	$\dfrac{1}{b-a}$ $a < y < b$	$\dfrac{e^{bt} - e^{at}}{t(b-a)}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Standard uniform $= U(0,1)$	1 $0 < y < 1$	$\dfrac{e^t - 1}{t}$	$1/2$	$1/12$
Normal $N(a,b^2)$	$\dfrac{1}{b\sqrt{2\pi}} e^{-\frac{1}{2b^2}(y-a)^2}$ $-\infty < y < \infty$	$e^{at + \frac{1}{2}b^2t^2}$	a	b^2
Standard normal $Z \sim N(0,1)$	$\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ $-\infty < y < \infty$	$e^{\frac{1}{2}t^2}$	0	1
Gamma $Gam(a,b)$	$\dfrac{y^{a-1}e^{-y/b}}{b^a\Gamma(a)}$ $y > 0$	$(1-bt)^{-a}$	ab	ab^2
Chi-square $\chi^2(n)$ $= Gam(n/2,2)$	$\dfrac{y^{\frac{n}{2}-1} e^{-y/2}}{2^{n/2}\Gamma(n/2)}$ $y > 0$	$(1-2t)^{-n/2}$	n	$2n$
Exponential $Expo(b)$ $= Gam(1,b)$	$\dfrac{1}{b}e^{-y/b}, \ y > 0$	$(1-bt)^{-1}$	b	b^2
Standard exponential $Expo(1)$	$e^{-y}, y > 0$	$(1-t)^{-1}$	1	1
Beta $Beta(a,b)$	$\dfrac{y^{a-1}(1-y)^{b-1}}{B(a,b)}$ $0 < y < 1$	no simple expression	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$

Mixed distributions & random variables

See Section 4·11 (Assessable)

: (TBC)

Defn Suppose a rv Y has cdf

$$F(y) = c\,F_1(y) + (1-c)\,F_2(y)$$

weighted average of F_1 & F_2

where: $0 < c < 1$

F_1 is the cdf of discrete rv X_1

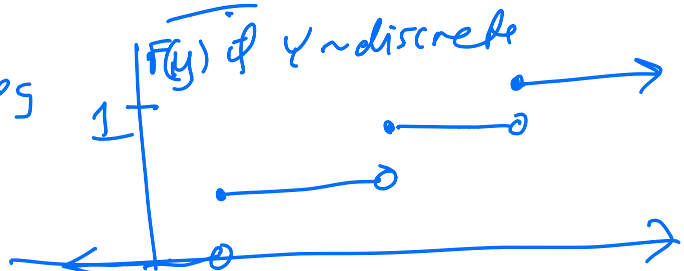
F_2 ------ continuous rv X_2.

Then we say that Y is a mixed rv & Y has a mixed dsn.

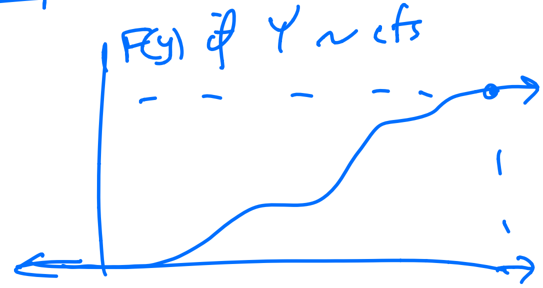
Note: If Y is mixed it is not cts & not discrete

The cdf of Y is like the cdf of a cts rv but it also has "jumps".

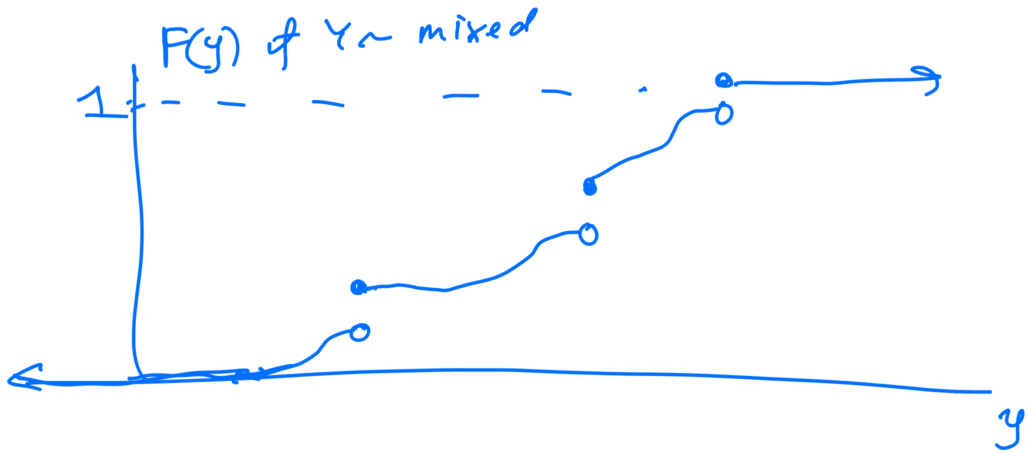
eg



piecewise constant (with jumps)



no jumps



Note: If $c = 1$ then $Y \sim$ discrete $(Y = X_1)$
$c = 0$ $Y \sim$ cts $(Y = X_2)$

Eg 1 2 coins are tossed.
If 2 Ts then ~~write~~ $Y = 0$
2 Hs $Y = 1$
O/w Y is a number chosen rand. & unif. between 0 & 1 (eg 0.6239)
Find the cdf, pdf & mean of Y.

Soln
$$\boxed{\begin{aligned} P(Y=0) &= P(TT) = 1/4 \\ P(Y=1) &= P(HH) = 1/4 \end{aligned}}$$
$$P(Y = 1.1) = 0$$

$P(Y \leq 0) = 1/4 = F(0)$
$P(Y \leq 1) = 1 = F(1)$
For $y < 0$, $P(Y \leq y) = 0 = F(y)$
For $y > 1$, $P(Y \leq y) = 1 = F(y)$

For $\underline{0<y<1}$

$F(y) = P(Y \le y) = ?$ (use LTP)

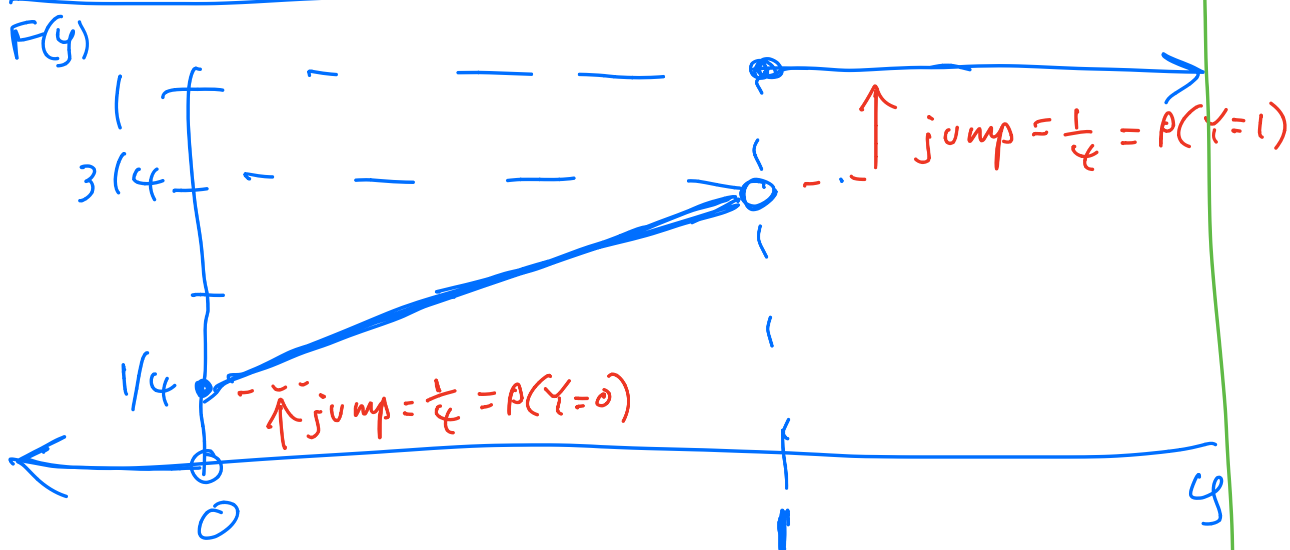
$$= P(TT)\,\underline{P(Y \le y \mid TT)} + P(HH)\,\underline{P(Y \le y \mid HH)} + P(HT \cup TH)\,\underline{P(Y \le y \mid HT \cup TH)}$$

$$= \frac{1}{4} P(Y=0 \mid TT) + \frac{1}{4}(0) + \frac{1}{2} P(Z \le y)$$

where $\underline{Z \sim U(0,1)}$

$$= \frac{1}{4} \times 1 + 0 + \frac{1}{2} y$$

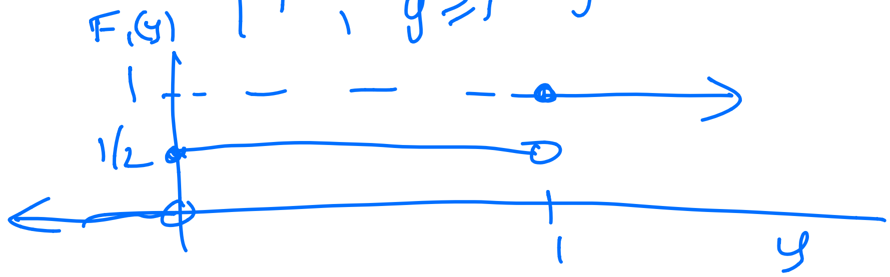
$$\boxed{\text{So } F(y) = \begin{cases} 0 & , \ y < 0 \\ \frac{1}{4} + \frac{1}{2} y & , \ 0 \le y < 1 \\ 1 & , \ y \ge 1 \end{cases}}$$

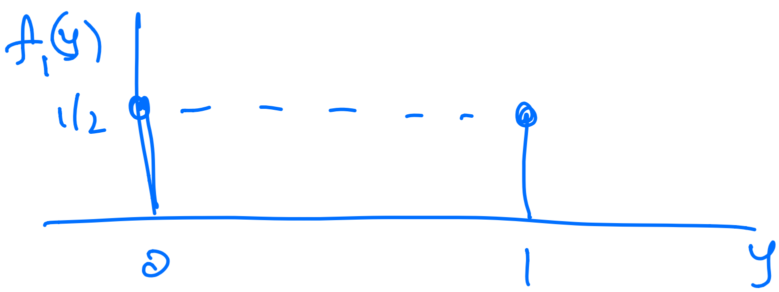


Note: $\boxed{F(y) = c\,F_1(y) + (1-c)\,F_2(y)}$

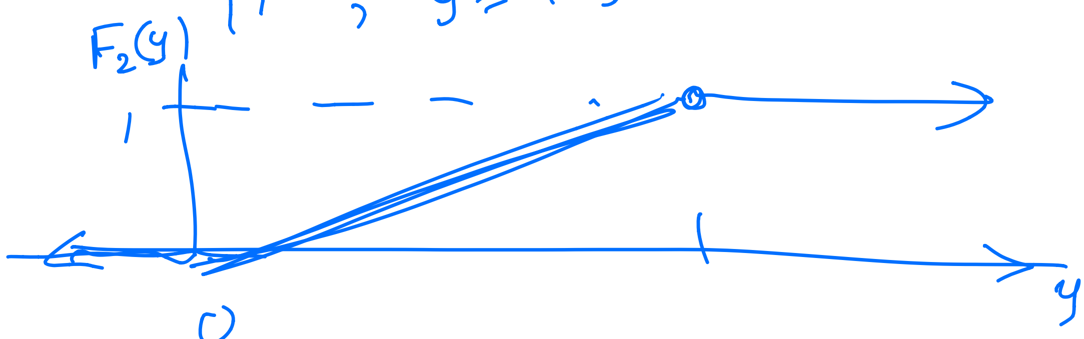
where: $c = 1/2$

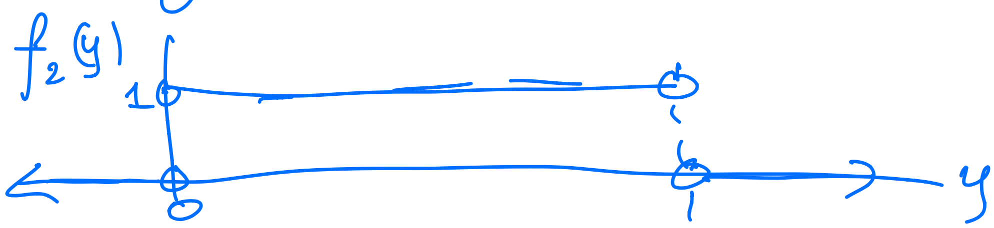
$$F_1(y) = \begin{cases} 0 & , \; y < 0 \\ 1/2 & , \; 0 \le y < 1 \\ 1 & , \; y \ge 1 \end{cases} = \text{cdf of } X_1 \sim \text{Bern}\left(\tfrac{1}{2}\right)$$





$$F_2(y) = \begin{cases} 0 & , \; y < 0 \\ y & , \; 0 \le y < 1 \\ 1 & , \; y \ge 1 \end{cases} = \text{cdf of } X_2 \sim U(0,1)$$





~~Ex~~ The pdf of Y is ? (not a probability & not a density)

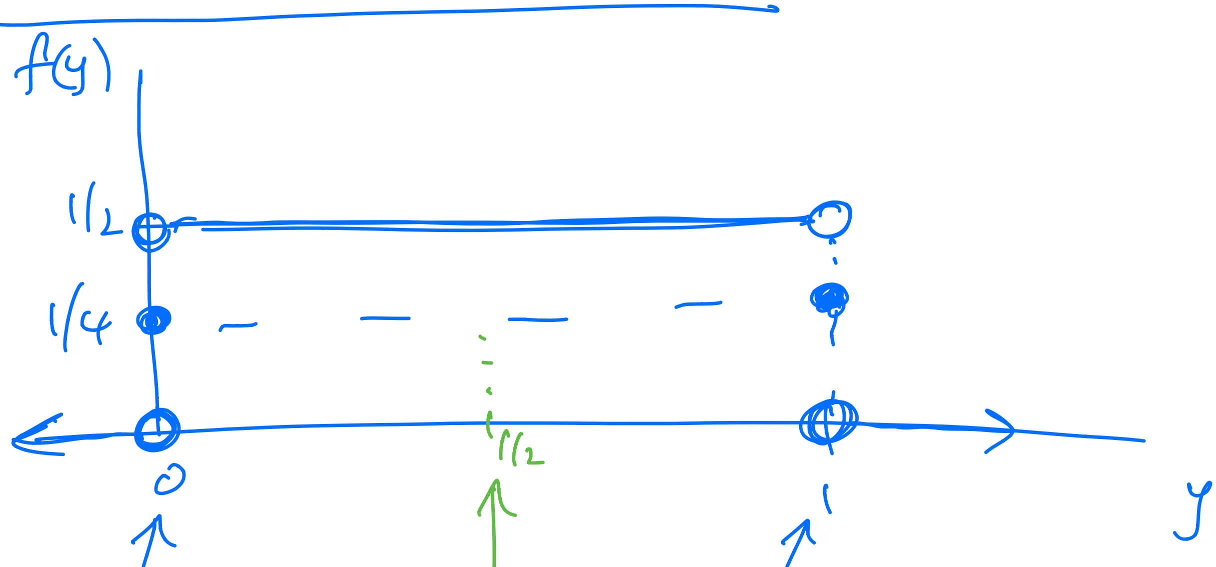
$$f(y) = \begin{cases} \text{jump at } 0, & y = 0 \\ \text{slope of } \left(\frac{1}{4} + \frac{1}{2}y\right) & 0 < y < 1 \\ \text{jump at } 1 & y = 1 \end{cases}$$

$$f(y) = \begin{cases} 1/4 & y = 0 \\ 1/2 & 0 < y < 1 \\ 1/4 & y = 1 \\ 0 & \text{o/w} \end{cases}$$

$$\boxed{f(y) = \begin{cases} 1/4, & y = 0, 1 \\ 1/2, & 0 < y < 1 \end{cases}}$$

← discrete part

← cts part



This is obviously EY

Note 1:

$$\boxed{\sum_{\text{discrete}} f(y) + \int_{\text{cts}} f(y)\,dy = 1}$$ (true for all mixed rvs Y)

Check: $LHS = [f(0) + f(1)] + \int_0^1 \frac{1}{2}\,dy$

$= [\frac{1}{4} + \frac{1}{4}] \quad [\frac{y}{2}|_0^1]$

$= \frac{1}{2} + \frac{1}{2} = 1$ (✓)

Note 2: We could write

$$\boxed{Y = RX_1 + (1-R)X_2}$$

where $R \sim Bern(1/2)$

and X_1, X_2 & $R \sim \perp$ (totally independent)

There's a 50% chance that $Y = X_1$ (if $R=1$)

" " " " " " " $Y = X_2$ (if $R=0$)

So Y is a function of rvs (see Ch6)

$$EY = \sum_{\text{discrete } y} y\,f(y) + \int_{\text{cts } y} y\,f(y)\,dy$$

$= [0 f(0) + 1 f(1)] + \int_0^1 y\left(\frac{1}{2}\right)dy$

$= [0 \times \frac{1}{4} + 1 \times \frac{1}{4}] + [\frac{y^2}{4}|_0^1]$

$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ (✓)

OR $EY = cEX_1 + (1-c)EX_2$

$= \frac{1}{2}\left(\frac{1}{2}\right) + \left(1-\frac{1}{2}\right)\frac{1}{2} = \frac{1}{2}$ (more easy)

Also $EY^2 = c\,EX_1^2 + (1-c)\,EX_2^2$

$= \dots$

$= 5/12$

So $VY = EY^2 - (EY)^2 = \dots = \frac{1}{6}$

See Sec. 4.11 & do some exercises there.