Submit HW1 before midnight. HW2 out due next Fri.
Reading 1.4 – 1.6 of Gallager's book & my notes lec2.pdf.

X: non-negative discrete rv. Sample values are $\{a_0, a_1, a_2, \dots\} \subset [0,\infty)$ $\quad \Pr(X \geq 0) = 1$

$$\mathbb{E}[X] = \sum_{j=0}^{\infty} a_j \, p_X(a_j)$$ where $p_X(\cdot)$ is the pmf of X.

We say that the expectation of X exists if $\mathbb{E}X < \infty$.

Rmk: Expectation need not exist if sample value is infinite.

Eg: $\Pr(X = n) = \frac{1}{n(n+1)}, \quad n \in \mathbb{N} = \{1, 2, \dots\}$

① Check $\sum_{n \in \mathbb{N}} \Pr(X = n) = 1$

② $\mathbb{E}X = \sum_{n \in \mathbb{N}} n \Pr(X = n) = \sum_{n \in \mathbb{N}} \frac{1}{n+1} = \infty$.

Fact: [Layer-Cake Representation]

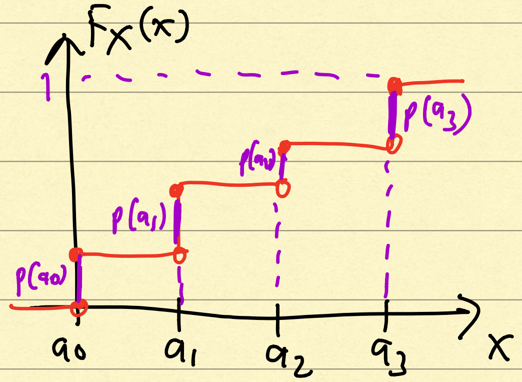
$$\mathbb{E}X = \int_0^{\infty} \underbrace{1 - F_X(x)}_{F_X^c(x)} \, dx$$ where $F_X(\cdot)$ is cdf of X.

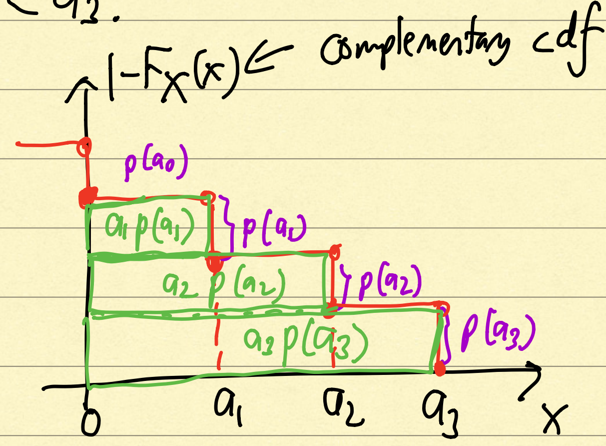
$F_X(x) = \Pr(X \leq x), \; x \in \mathbb{R}$

Pf: What does the cdf look like for a nonneg.

discrete rv $X \in \{a_0, a_1, a_2, a_3\} \subset \mathbb{R}_+$?
$a_0 = 0$

Say $a_0 = 0 < a_1 < a_2 < a_3$.





$$\int_0^\infty 1 - F_X(x)\,dx = \text{Area under the comp. cdf } [0,\infty)$$

$$\mathbb{E}X = a_1 p(a_1) + a_2 p(a_2) + a_3 p(a_3) = \sum_{i=0}^{3} a_i p_X(a_i)$$

Fact: For a non-negative integer-valued rv X,

$$\mathbb{E}X \overset{\checkmark}{=} \sum_{n=0}^{\infty} P(X>n) = \sum_{n=1}^{\infty} P(X \geq n).$$

Pf: $\sum_{n=0}^{\infty} P(X>n) = Pr(X>0) + Pr(X>1) + Pr(X>2) + \dots$

$= [p(1) + p(2) + p(3) + \dots] +$

$[0 + p(2) + p(3) + \dots] +$

$[\ \dots\ ]$

$[\,0 + 0 + p(3) + p(4) + \cdots\,]$

$$= 1\cdot p(1) + 2\cdot p(2) + 3\cdot p(3) + \cdots$$

$$= \sum_{n=1}^{\infty} n\, p_X(n) = \mathbb{E}X.$$

$$\shortparallel$$

$$\sum_{n=0}^{\infty} n\, p_X(n)$$

Fact: X cts rv & $\Pr(X \geq 0) = 1$. (non-neg.)

$$\mathbb{E}X = \int_0^{\infty} x f_X(x)\,dx = \int_0^{\infty} F_X^c(x)\,dx \qquad = 1 - F_X(x)$$

Pf: $\text{RHS} = \int_0^{\infty} \Pr(X > x)\,dx$

$$= \int_0^{\infty} \left(\int_x^{\infty} f_X(u)\,du \right) dx$$

indicator $f^{\underline{n}}$

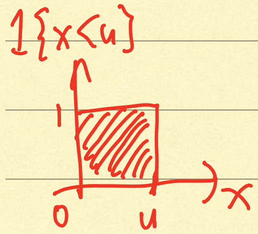
$$= \int_{x=0}^{\infty} \int_{u=0}^{\infty} f_X(u)\, \mathbb{1}\{u > x\}\,du\,dx.$$

Tonelli

Fubini

const wrt x

$$= \int_{u=0}^{\infty} \int_{x=0}^{\infty} f_X(u)\, \mathbb{1}\{u > x\}\,dx\,du$$



$$= \int_{u=0}^{\infty} f_X(u) \underbrace{\left(\int_{x=0}^{\infty} \mathbb{1}\{u > x\}\,dx \right)}_{u} du$$

$$= \int_0^{\infty} f_X(u)\,u\,du = \text{LHS}. \qquad ///$$

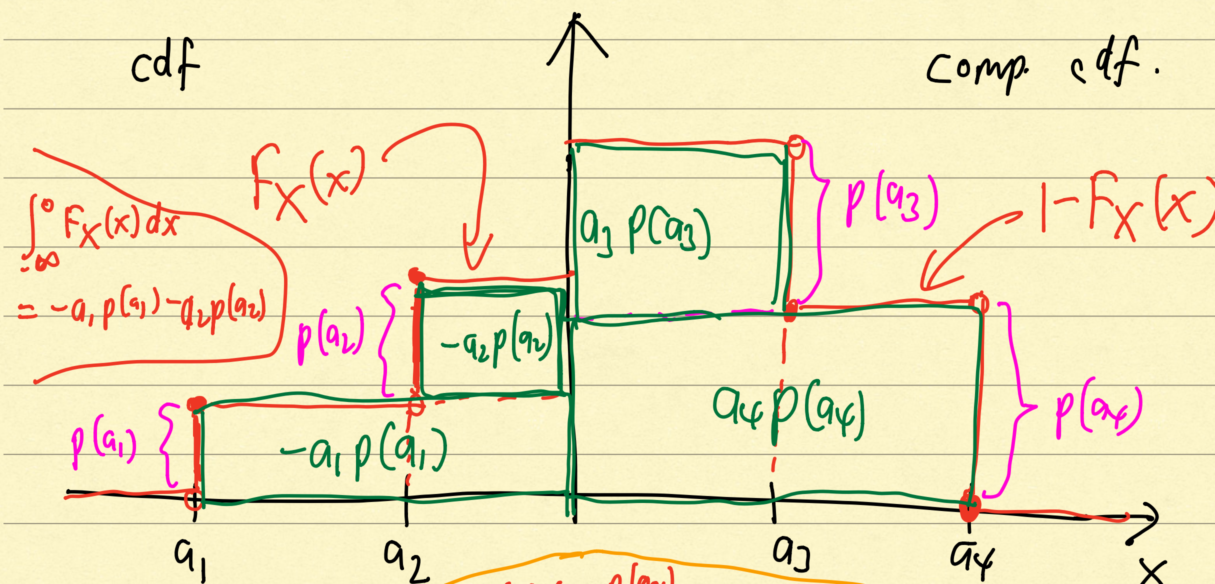
X: has both positive & neg. sample values.

$X \in \{a_1, a_2, \dots\} \subset \mathbb{R}$.

$$\mathbb{E}X = \sum_i a_i \, p_X(a_i) = \sum_{i: a_i \le 0} a_i p(a_i) + \sum_{i: a_i > 0} a_i p(a_i)$$

Can be expressed in term $F_X(x)$ (cdf of X).

$$a_1 < a_2 < 0 < a_3 < a_4.$$



$$\mathbb{E}X = \underbrace{\int_{-\infty}^{0} - F_X(x)\,dx}_{= a_1 p(a_1) + a_2 p(a_2)} + \underbrace{\int_0^{\infty} 1 - F_X(x)\,dx}_{= a_3 p(a_3) + a_4 p(a_4)} \quad -(*)$$

$$= \int_{-\infty}^{\infty} \left(u(x) - F_X(x)\right) dx. \quad \Bigg| \; u(x) = \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

Def: For a general rv X, $\underline{\mathbb{E}X \text{ exists}}$ with the value given in $(*)$ if each of the 2

terms is finite.
↳ (integrals)

RVs as functions of other rvs.

X: rv $\qquad h: \mathbb{R} \to \mathbb{R}$.

Define $\quad Y = h(X)$.

Fact: $\qquad$ find the pdf of Y

$$\mathbb{E}Y = \int_{\mathbb{R}} y f_Y(y)\, dy$$

$$= \int_{\mathbb{R}} h(x) f_X(x)\, dx$$

Some f^{ns} are more impt than others.

$h(x) = x^k \qquad \mathbb{E}Y = \int x^k f_X(x)\, dx$: k^{th} moment of X.

$h(x) = (x - \bar{x})^2$, $\quad \bar{x} = \mathbb{E}X$; expectation

$$\mathbb{E}Y = \int_{\mathbb{R}} h(x) f_X(x)\, dx = \int_{\mathbb{R}} (x - \bar{x})^2 f_X(x)\, dx$$

$$= \mathbb{E}[(X - \bar{x})^2]$$: Variance of X.

$$\mathrm{Var}(X) \neq \mathbb{E}(X^2)$$

Ex: $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}[X])^2$.

If $\bar{x} = 0$, $Var(X) = \mathbb{E}(X^2)$

The value of x that achieves $\min_{x \in \mathbb{R}} \underbrace{\mathbb{E}[(X-x)^2]}_{=g(x)}$

is $\bar{x} = \mathbb{E}(X)$.

$$g'(x) = \frac{d}{dx} \mathbb{E}[(X-x)^2]$$

$$= \mathbb{E}[\cancel{2}(X-x)] = 0$$

Connection btw mean & variance.

$$x = \mathbb{E}X.$$

Another f^n of interest is $h(x,y) = x+y$.

$$Z = h(X,Y) = X+Y.$$

If $X \perp\!\!\!\perp Y$, we can express the $\underbrace{\text{dist}^n}_{\text{pdf.}}$ of Z in terms of the dist^n of X & Y.

$$\underbrace{F_Z(z)}_{=\text{cdf of } Z} = Pr(Z \leq z)$$

$$= Pr(\underbrace{X+Y}_{=Z} \leq z)$$

law of total prob.

$$= \int_{-\infty}^{\infty} f_X(x) Pr(X + Y \leq z \mid X = x)\, dx$$

$$= \int_{-\infty}^{\infty} f_X(x) Pr(x + Y \leq z \mid \cancel{X = x})\, dx.$$

$$= \int_{-\infty}^{\infty} f_X(x)\, \Pr(x + Y \le z)\, dx \quad [X \perp\!\!\!\perp Y]$$

$$= \int_{-\infty}^{\infty} f_X(x)\, \Pr(Y \le z - x)\, dx$$

$$= \int_{-\infty}^{\infty} f_X(x)\, F_Y(z-x)\, dx.$$

We have shown:

$$F_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, F_Y(z-x)\, dx.$$

Differentiate both sides wrt z

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz}\left(\int_{-\infty}^{\infty} f_X(x)\, F_Y(z-x)\, dx\right)$$

pdf of Z

$$= \int_{-\infty}^{\infty} f_X(x) \left(\frac{d}{dz} F_Y(z-x)\right) dx$$

$$= \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

Thm: If X & Y are <u>indep</u>, then the pdf of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

$$\parallel$$

Convolution of f_X & f_Y.

Fact: For arbitrary rvs $X_1, \dots, X_n$

$$\mathbb{E}[X_1 + \dots + X_n] = \mathbb{E}X_1 + \dots + \mathbb{E}X_n.$$

Linearity of expectation. (Exciting Thm later on)

Fact: For $X_1, \dots, X_n$ independent (actually uncorrelatedness is enough $\mathbb{E}[X_i X_j] = (\mathbb{E}X_i)(\mathbb{E}X_j)$ for $i \neq j$), then

$$\mathrm{Var}\left(\sum_{i=1}^n \alpha_i X_i\right) = \sum_{i=1}^n \alpha_i^2 \mathrm{Var}(X_i).$$

Pf sketch: $\alpha_1 = \alpha_2 = 1$, $\mathbb{E}X_1 = \mathbb{E}X_2 = 0$

$$\begin{aligned}
\mathrm{Var}(X_1 + X_2) &= \mathbb{E}[(X_1 + X_2)^2]. \\
&= \mathbb{E}(X_1^2) + 2\mathbb{E}[X_1 X_2] + \mathbb{E}(X_2^2) \\
&= \mathbb{E}(X_1^2) + 2\cancel{\mathbb{E}X_1}^{0}\cancel{\mathbb{E}X_2}^{0} + \mathbb{E}(X_2^2) \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2)
\end{aligned}$$

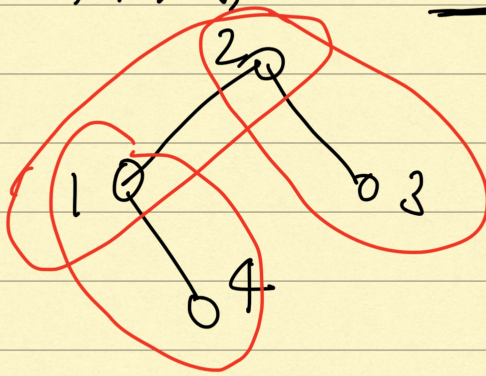
///

Example of the utility of Linearity of expectation.

"The Probabilistic Method" (Alon & Spencer).

Def: A graph $G = (V, E)$ consists of a vertex set (node set) $V = \{1,2,\dots,n\}$ and a set of pairs of vertices $E \subset \binom{V}{2}$

$V = \{1,2,3,4\}$, $E = \{\{1,2\}, \{2,3\}, \{1,4\}\}$.



Def: A hypergraph $H = (V,E)$ consists of a set of vertices $V = \{1,\dots,n\}$ and a set of tuples of vertices E. Each $e \in E$ is called a hyperedge.

Eg: If $n = 10$, $V = \{1, 2, \dots, 10\}$

$E = \{ \{1,2,3\}, \{2,5\}, \{4,5,6\}, \{3,4,7\}, \{8,9,10,7\}\}$

Def: A k-uniform hypergraph $H = (V,E)$ is a hypergraph if each $e \in E$ contains exactly k nodes.

Eg: If $n = 10$, $V = \{1, 2, \dots, 10\}$

$E = \{ \overbrace{\{1,2,3\}}^{\text{not-mc}}, \overbrace{\{2,5,10\}}^{\text{not mc}}, \{4,5,6\}, \{3,4,7\}, \{8,9,10\} \}$

This is a 3-uniform hypergraph (3-regular)

Def: An edge e in H is <u>monochromatic</u> if all the nodes in e have the same color

Def: A hypergraph H is <u>2-colorable</u> if we can color all nodes in V so that no edge is monochromatic

Rmk: If H has more hyperedges, it is <u>less</u> likely to be 2-colorable.

<u>Thm [Erdős (1963)]</u>: Every k-uniform hypergraph with $\underline{\underline{< 2^{k-1}}}$ edges is 2-colorable.

Rmk: Erdős (1964) showed $\exists$ k-uniform hypergraph with $O(k^2 2^k)$ hyperedges that is not 2-colorable.

Pf: Let $H = (V,E)$ be a k-unif. hypergraph w/ $< 2^{k-1}$ edges.

Color each $v \in V$ with one of 2 colors (blue or red) with eq. prob.

$\forall e \in E$,

Define the rv $X_e = \begin{cases} 1 & e \text{ is monochromatic.} \\ 0 & \text{else} \end{cases}$

Consider $X = \sum_{e \in E} X_e$ total # of monochromatic edges in H.

$\mathbb{E} X_e = \cancel{1} \cdot \Pr(e \text{ is monochromatic}) + \cancel{0 \cdot \Pr(e \text{ not MC})}$. $\to 0$.

$= \Pr(e \text{ is all blue}) + \Pr(e \text{ is all red})$

$= 1/2^k + 1/2^k = 2^{-k+1}$.

Hence, by linearity of expectation,

$$\mathbb{E} X = \sum_{e \in E} \boxed{\underbrace{\mathbb{E} X_e}_{= 2^{-k+1}}} = (\text{num of edges in } H) \cdot \boxed{2^{-k+1}}$$

$$< 2^{k-1} \cdot 2^{-k+1} = 1.$$

$X \in \{0, 1, \cdots\}$

In conclusion $\mathbb{E} X < 1$

But $\mathbb{E} X = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2\, p_X(2) + \cdots$

Since $\mathbb{E} X < 1$, $p_X(0) > 0$.

(Why? Suppose, to the contrary, $p_X(0) = 0$, then
$\mathbb{E} X = 1 \cdot p_X(1) + 2 p_X(2) + \cdots \geq 1$. $\because \sum_{x \geq 1} p_X(x) = 1$)

$p_X(0) = P(X = 0) > 0$ means that the prob. that <u>H has 0 monochromatic edges is positive.</u>

Hence the prob. that H is 2-colorable is positive. ////

Conditional Expectation

X, Y are discrete rvs w/ $p_Y(y) > 0 \quad \forall y$.
$$p_Y(y) = \Pr(Y=y)$$

Conditional expectation of X given event $\{Y=y\}$.

$$\mathbb{E}[X \mid Y=y] = \sum_x x \, p_{X|Y}(x|y)$$

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)} \quad \text{exists } \forall x, y.$$

Conditional expectation of X given rv Y.

$$g(y) := \mathbb{E}[X \mid Y=y] \in \mathbb{R}.$$

$g(y)$ is a f^n of y

Consider $g(Y)$ where Y follows the dist^n p_Y.

Def: The conditional expectation of rv X given rv Y is $\mathbb{E}[X|Y] = \underline{g(Y)}$.

Note: $\mathbb{E}[X|Y]$ is a f^n of Y.

Fact: [Law of iterated expectation / Tower property].

$$\mathbb{E}[\underbrace{\mathbb{E}[X|Y]}_{\text{function of } Y}] = \mathbb{E}X.$$

Pf: $LHS = \mathbb{E}_Y[\underbrace{g(Y)}_{\mathbb{E}[X|Y]}] = \sum_y g(y)\, p_Y(y)$

$= \sum_y \mathbb{E}[X|Y=y]\, p_Y(y)$

$= \sum_y \left(\sum_x x\, p_{X|Y}(x|y)\right) p_Y(y)$

$= \sum_x x \sum_y p_{X|Y}(x|y)\, p_Y(y)$

$= \sum_x x \left(\sum_y p_{X,Y}(x,y)\right)$

$= \sum_x x\, p_X(x) = \mathbb{E}X = RHS.$ as desired ////.

Ex: X_1 : outcome of the first toss of a 6-sided dice

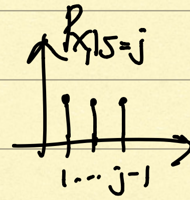
$\perp\!\!\!\perp$

X_2 : outcome of the second toss of a 6-sided dice

$S = X_1 + X_2 \in \{2, 3, \ldots, 11, 12\}.$

Condition on event $S=j \in \{2, 3, \dots, 7\}$.

$$X_1 \sim \text{Unif}\{1, 2, \dots, j-1\}.$$



Condition on event $S=j \in \{8, 9, \dots, 12\}$.

$$X_1 \sim \text{Unif}\{j-6, \dots, 6\}$$

(e.g.) if $S=9$, $X_1 \sim \text{Unif}\{3, \dots, 6\}$).

$$\mathbb{E}[X_1 \mid S=j] = \begin{cases} \frac{1+(j-1)}{2}, & 2 \le j \le 7 \\ \frac{(j-6)+6}{2} & 8 \le j \le 12 \end{cases}$$

$$\mathbb{E}[X_1 \mid S=j] = \begin{cases} j/2, & 2 \le j \le 7 \\ j/2, & 8 \le j \le 12 \end{cases}$$

$$\mathbb{E}[X_1 \mid S=j] = j/2, \quad 2 \le j \le 12.$$

Thus, $Y = \mathbb{E}[X_1 \mid S]$ is a discrete rv that is a function of S takes on values $\frac{2}{2}, \frac{3}{2}, \frac{4}{2}, \dots, \frac{11}{2}, \frac{12}{2}$. ($j/2$ for $2 \le j \le 12$)

$$\boxed{Y = \frac{S}{2}}$$

$$p_Y(j/2) = p_S(j), \quad j = 2, 3, \dots, 12.$$

$$\mathbb{E}[X_1] = \mathbb{E}_S[\mathbb{E}[X_1 \mid S]] \overset{\text{Exp.}}{=} \sum_{j=2}^{12} j/2 \, \underline{p_Y(j/2)}$$

$$= \sum_{j=2}^{12} j/2 \, p_S(j) = \frac{1}{2}\mathbb{E}[S] = 7/2.$$

Moment Generating Function (MGF).

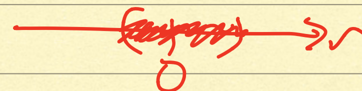
X: r.v. $\qquad h(x) = e^{rx}, \quad r \in \mathbb{R}.$

$$E[h(X)] = g_X(r) = E[e^{rX}] = \int_{-\infty}^{\infty} e^{rx} f_X(x)\,dx.$$

pdf

$$g_X(r) = \sum_x e^{rx} p_X(x)$$

pmf



Cumulant generating $f^{\underline{n}}$: $\ln g_X(r)$.

If $g_X(r)$ exists in a neighborhood around 0, then its derivatives of all orders exist in that nbd.

$$\frac{d}{dr} g_X(r) = \frac{d}{dr} \int_{-\infty}^{\infty} e^{rx} f_X(x)\,dx.$$

$$= \int_{-\infty}^{\infty} \frac{d}{dr} e^{rx} f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} x e^{rx} f_X(x)\,dx.$$

$$\left.\frac{d}{dr} g_X(r)\right|_{r=0} = \int_{-\infty}^{\infty} x f_X(x)\,dx = EX.$$

More generally,

$$\frac{d^k}{dr^k} g_X(r)\Big|_{r=0} = \mathbb{E}(X^k) \quad k^{th} \text{ moment of } X.$$

Fact: $X_1, \cdots, X_n$ mutually Indep.

$$S_n = X_1 + \cdots + X_n.$$

$$g_{S_n}(r) = \mathbb{E}[e^{rS_n}] = \mathbb{E}[e^{r(X_1+\cdots+X_n)}]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n} e^{rX_i}\right]$$

$$= \prod_{i=1}^{n} \mathbb{E}[e^{rX_i}] = \prod_{i=1}^{n} g_{X_i}(r)$$

If $\{X_i\}$ are i.i.d. with common dist. f_X,

$$g_{S_n}(r) = (g_X(r))^n.$$

Def: The k^{th} moment of X is $\mathbb{E}[X^k]$.

The k^{th} absolute moment of X is $\mathbb{E}[|X|^k]$

Def: The characteristic f^n of X is

$$h(s) = \mathbb{E}[e^{isX}].$$

If $g : \mathbb{R} \to \mathbb{R}$ is strictly monotonically $\uparrow$ & g^{-1} diff$^{\text{ble}}$,

$$\mathbb{E}[g(X)] = \int g(x) f_X(x)\,dx = \int y f_Y(y)\,dy$$

where $Y = g(X)$. The generalization to general g is left as a tedious ex.

Pf: Let's find $f_Y(y)$ in terms of $f_X(x)$.

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) \overset{\text{strict mon.}\uparrow}{=} P(X \le g^{-1}(y))$$
$$= F_X(g^{-1}(y)).$$

$$\Rightarrow f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy}\left(g^{-1}(y)\right)$$
$$= f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}.$$

$$\int y f_Y(y)\,dy = \int y f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}\,dy$$

$$\underset{x = g^{-1}(y)}{=} \int g(x) f_X(x) \frac{1}{\cancel{g'(x)}} \cancel{g'(x)}\,dx$$

$y = g(x)$

$dy = g'(x)\,dx$

$$= \int g(x) f_X(x)\,dx. \quad ////.$$