Similarly, the support of $F_T$ is $\mathbb{R}^+$, and its marginal pdf is

$$f_T(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dx = \int_{-\infty}^{\infty} \frac{21}{4} x^2 y \, dx$$

$$= \begin{cases} \frac{7}{2} y^3 & \text{for } 0 < y < 1 \\ 0 & \text{else} \end{cases}$$

If you have continued the joint dist. $f_{X,Y}(x,y)$, you can reconstruct the marginals $f_X(x)$ and $f_Y(y)$, but not the other way around: if all you have is the marginals, they do not uniquely determine the joint.
Example

Case 1: \( X = \# \) heads in \( n \) tosses of fair coin 1 and independently \( Y = \# \) heads in \( n \) tosses of fair coin 2.

\( X = \# \) heads in \( n \) tosses of fair coin 1

\( Y = X \)

\( X \sim \text{Binomial} \left( n, \frac{1}{2} \right) \)

so \( f_{X}(x) = \binom{n}{x} \left( \frac{1}{2} \right)^{x} \left( 1 - \frac{1}{2} \right)^{n-x} \quad x = 0, 1, \ldots, n \)

is also

and \( Y \sim \text{Binomial} \left( n, \frac{1}{2} \right) \)

so \( f_{Y}(y) = \binom{n}{y} \left( \frac{1}{2} \right)^{n} \quad y = 0, 1, \ldots, n \)

Since \( X \) and \( Y \) are independent in

Case 1, \( f_{X,Y}(x,y) = f_{X}(x) \cdot f_{Y}(y) \)

(as we'll see in a minute).
So in case 1:
\[ f_{X,Y}(x,y) = \begin{cases} \binom{n}{x} \left( \frac{1}{2} \right)^n & \text{for } x = 0, 1, \ldots, n \\
0 & \text{else} \end{cases} \]

However: In case 2, \( X \) is Binomial \( (n, \frac{1}{2}) \) and so is \( Y \) (same as in case 1), but their joint distribution (since \( Y = X \)) is:
\[ f_{X,Y}(x,y) = \begin{cases} \binom{n}{x} \left( \frac{1}{2} \right)^n & \text{for } x = y = 0, \ldots, n \\
0 & \text{else} \end{cases} \]

Here is one situation in which the marginals uniquely determine the joint: when \( X \) and \( Y \) are independent.
Def. rvs \( X \) and \( Y \) are independent (non-vice) if for every sets \( A \) and \( B \) of real numbers \( P(X \in A \text{ and } Y \in B) = P(X \in A) \cdot P(Y \in B) \).

Consequence:
1. Immediately you get that if \( X \) and \( Y \) are independent

\[
F_X(x,y) = P(X \leq x \text{ and } Y \leq y)
\]

\[
= P(X \leq x) \cdot P(Y \leq y)
\]

\[
= F_X(x) \cdot F_Y(y).
\]

This is an if-then converse is also true

2. Differentiate this equation once with respect to \( x \) and once with respect to \( y \) to get the result that

\[
\frac{\partial^2}{\partial x \partial y} f_{X,Y}(x,y) = \frac{\partial}{\partial x} f_X(x) \cdot \frac{\partial}{\partial y} f_Y(y).
\]

\[
\frac{\partial^2}{\partial x \partial y} f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).
\]
Example Suppose that continuous rvs \( X \) and \( Y \) have joint pdf

\[
f_{XY}(x, y) = \begin{cases} \frac{1}{24} x^2 y^2 & \text{for } x^2 + y^2 \leq 1 \\ 0 & \text{else} \end{cases}
\]

The support \( S_{XY} \) of \( f_{XY} \) is the region inside the unit circle. You can evaluate the normalizing constant by computing \( \iint_{S_{XY}} k x^2 y^2 \, dx \, dy \) and setting it equal to 1:

\[
1 = \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} k x^2 y^2 \, dy \, dx = \frac{\pi}{24}
\]

So \( k = \frac{24}{\pi} \)

\[\square\] Are \( X \) and \( Y \) independent?
A: No, they can’t be: since the only points with positive density satisfy 
\( x^2 + y^2 \leq 1 \), for any given value of \( y \), the possible value of \( x \) depends on \( y \), & vice versa. Example:

Continuous rv \( X \) and \( Y \) have joint pdf

\[
f_{XY}(x,y) = \begin{cases} 
  k e^{-(x+y)} & \text{for } x \geq 0 \text{ and } y \geq 0 \\
  0 & \text{else}
\end{cases}
\]

Q: Are \( X \) and \( Y \) independent?

A: Yes, because (a) \( e^{-(x+y)} \) factors into \((e^{-x})(e^{-y})\) and (b) the support \( \mathbb{S}_{xy} \) also ‘factors’: \((x \geq 0)A\,(y \geq 0)\)
Just choose \((k, k_x, k_y)\) such that
\[
\int_0^\infty k e^{-(x+2y)} \, dx \, dy = 1, \quad \int_0^\infty k_x e^{-x} \, dx = 1,
\]
\[
\int_0^\infty k_y e^{-2y} \, dy = 1, \quad \text{and} \quad k = k_x \cdot k_y.
\]

You get \(k_x = 1\), \(k_y = 2\), \(k = 2\). \(\checkmark\)

\[\text{Conditional probability distributions (as long as } P(A) > 0)\text{, we should be able to extend this idea to random variables.}\]

\[\text{Start with } X \text{ and } Y \text{ both discrete, so that we can talk about } P(Y = y | X = x)\]
Def. If $X$ and $Y$ have a discrete joint distribution with joint pdf $f_{X,Y}(x,y)$ and $X$ has marginal pdf $f_X(x)$, then for each $x$ such that $f_X(x) > 0$ define

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

to be the conditional pdf of $Y$ given $X$.

Example:

gender & marijuana legalization preference at UCLA

(See doc. comm. notes)

Quiz 3
Now let's do the analogous thing for continuous rvs.

**Def.** If $X$ and $Y$ have a continuous joint distribution with joint pdf $f_{X,Y}(x,y)$ and $X$ (continuous) has marginal pdf $f_X(x)$, then for each $x$ such that $f_X(x) > 0$, define

$$f_{Y|X}(y|x) = \left\{ \begin{array}{ll} \frac{f_{X,Y}(x,y)}{f_X(x)} & \text{to be} \\ \end{array} \right. $$

the conditional pdf of $Y$ given $X$.

**Continuity or earlier example**

If $X, Y$ have joint pdf

$$f_{X,Y}(x,y) = \left\{ \begin{array}{ll} \frac{24}{5} xy & \text{for} \: x, y \in (0, 5) \\ 0 & \text{else} \end{array} \right. $$
let's work out \( f_{\text{X|Y}}(y|x) \) and

\[
f_{\text{X|Y}}(x|y) = \begin{cases} \frac{21}{8} x^2 (1-x^4) & \text{for } -1 \leq x \leq 1 \\ 0 & \text{else} \end{cases}
\]

\[
f_Y(y) = \begin{cases} \frac{7}{8} y^{5/2} & \text{for } 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}
\]

Immediately, then, for all \( x \) for which \( f_{\text{X|Y}}(x,y) > 0 \), namely \(-1 < x < 1\)

\[
f_{\text{X|Y}}(y|x) = \frac{f_{\text{X|Y}}(x,y)}{f_X(x)}
\]

\[
= \begin{cases} \frac{21}{4} x^2 y & \text{for } 0 \leq x^2 \leq y \leq 1 \\ \frac{21}{8} x^2 (1-x^4) & \text{else} \end{cases}
\]
A few points:

1. \( f(1, y) = \lim_{x \to \frac{3}{2}} \frac{y}{x} = \frac{y}{\frac{3}{2}} = \frac{2y}{3} \)

2. For a limit, \( f(x, y) = \lim_{x \to 1} (\frac{y}{x}) \)

But note that what's actually being computed when \( X \) and \( Y \) are continuous:

\[ f(1) = \lim_{x \to 1} f(x, 1) = \lim_{x \to 1} \frac{1}{x} = 1 \]

Note:

- \( x = y \) for \( X \) is defined to involve
- \( y = 1 \) for \( Y \) does not have a limit as \( X \) goes to 1.

\( f(x, y) = \frac{y}{x} \) as \( x \to 1 \).
of $x$ values of width $\varepsilon$ around $X = x^*$.

Compute $P(\exists y | X \text{ is in the strip})$,

differentiate the result with respect to $y$,
and let $\varepsilon \to 0$. Thus you can
think of $f_{2|1|2}(y | x)$ as the conditional
pdf of $Y$ given that $X$ is close to $x$.

Constructing
a joint pdf
from marginals
& conditionals

we know that (as long
as no division by 0
happen)

$$f_{2|1|2}(y | x) = \frac{f_{2|2}(x, y)}{f_2(x)}$$

and

$$f_{1|2}(x | y) = \frac{f_{1|2}(x, y)}{f_1(y)}.$$
Multiply equation 1 by $f_\theta(x)$ and equation 2 by $f_\theta(\gamma)$ to get

$$f_{X \mid Y}(x, \gamma) = f_\theta(x) f_{X \mid Y \mid \gamma}(\gamma \mid x)$$

$$= f_\theta(\gamma) f_{X \mid \gamma}(x \mid \gamma).$$

So there are two ways to construct a joint pdf from marginal pdfs and a conditional pdf.

A machine produces nuts and bolts, and the nut paired with a particular bolt in the manufacturing process is
supposed to fit snugly on the bolt, let’s call a (nut, bolt) pair defective if the correct snug fit doesn’t happen (e.g., bolt diameter either too big or too small, or nut diameter too small or too big). Let \( \Theta = \frac{\text{proportion of defective bolts if the machine were allowed to run for an indefinitely long period}}{\text{time interval}}, \) \( \Theta \) is unknown.

Since we can only observe the machine for a finite (short) time interval, \( \Theta \) is unknown.

To learn about \( \Theta \), we could take a random sample of (nut, bolt) pairs of size \( m \) (say) and

Implicit assumption (stationarity): \( \Theta \) is constant over the entire indefinite time period.
Count the number of defectives in the sample (call this \( N \)).

\[
\frac{1}{\Pi^c \text{ stationary}}
\]

\[
(D_i | \theta) \sim \text{Bernoulli}(\theta)
\]

\[
(D_i | \theta) = \begin{cases} 1 & \text{if (but bolt) i} \text{ is defective} \\ 0 & \text{else} \end{cases}
\]

\[
N = \sum_{i=1}^{m} D_i
\]

so the total \( N \) is fixed & known

\[
f_N(n | m, \theta) = \binom{m}{n} \theta^n (1-\theta)^{m-n}
\]

Suppose that

\[
m = 114, \; N = 3
\]

A reasonable estimate of \( \theta \) would be

\[
\hat{\theta} = \frac{N}{m} = \frac{3}{114} = 2.66\%
\]

but how much uncertainty do we have about \( \theta \) on the basis of this dataset?
Bayesian story: vector \( \mathbf{D} = (D_1, \ldots, D_n) \) dataset

- Probability: \( p(\text{data} | \text{unknown}) \) easy
- \( p(\Theta | N, \Theta) = \ast \)
- \( p(D | \Theta) = p(D | \Theta) \)

Bayes' theorem

\[
p(\Theta | D) = \frac{p(D | \Theta)p(\Theta)}{p(D)}
\]

Total info about \( \Theta \) from dataset external to dataset

Info about \( \Theta \) internal to dataset

Because Bernoulli dataset \( D = (D_1, \ldots, D_n) \) and the rv \( N \) carry the same info about \( \Theta \)
Multivariate distributions So far we've looked at one and then two rvs at a time; easy to generalize to a finite number of rvs $\Xi_1, \ldots, \Xi_n$, a positive finite integer.

Def. The joint CDF of $n$ rvs $\Xi_1, \ldots, \Xi_n$ is the function $F_{\Xi_1, \ldots, \Xi_n}(\gamma_1, \ldots, \gamma_n)$ specified by $F_{\Xi_1, \ldots, \Xi_n}(\gamma_1, \ldots, \gamma_n) = P(\Xi_1 \leq \gamma_1, \ldots, \Xi_n \leq \gamma_n)$.

More compact to use vector notation: $\tilde{\Xi} = (\Xi_1, \ldots, \Xi_n)$, $\gamma = (\gamma_1, \ldots, \gamma_n)$.

$F_{\tilde{\Xi}}(\gamma) = P(\Xi_1 \leq \gamma_1, \ldots, \Xi_n \leq \gamma_n)$ is said to be a random vector taking values in $\mathbb{R}^n$. 
Def. If rv $(Z_1, \ldots, Z_n)^T$ have a discrete joint distribution if the random vector $Z$ can only take on a finite or countably infinite # of possible values $(y_1, \ldots, y_m) \in \mathbb{R}^n$.

The joint mass probability function of $Z$ is $f_{Z}(y_1, \ldots, y_m) = P(Z_1 = y_1, \ldots, Z_n = y_m)$ or equivalently $f_{Z}(y) = P(Z = y)$.

Example: in patients in treatment group randomized of a clinical trial; $\theta_i = \{ 1 \text{ if patient i has a good outcome} \} \text{ or else}$.

If nothing else is known about the patients (e.g., age, disease burden at start of trial, ...) it would be reasonable to model the $\theta_i$ as IID Bernoulli ($\theta$) success probability.
If $\theta$ were known, you could use $f_{\theta}(\mathbf{b})$ to predict the dataset before it arrives. By the IID assumption,
\[
P(\mathbf{b}_1 = b_1, \ldots, \mathbf{b}_n = b_n) = \prod_{i=1}^n P(\mathbf{b}_i = b_i) = \prod_{i=1}^n \theta^{b_i} (1 - \theta)^{1 - b_i}
\]

Recall that
\[
P(\mathbf{b}_i = b_i) = \theta^{b_i} (1 - \theta)^{1 - b_i} \text{ for } b_i = 0, 1
\]

Define $n \equiv n_1, \ldots, n_n$ have a continuous joint distribution if you can find a function $f_n$ on $\mathbb{R}^n$ such that
\[
f_n(\mathbf{s}) = \prod_{i=1}^n f_i(s_i)
\]

with $s = \sum_{i=1}^n s_i$. If $f_i$ is continuous, the joint distribution is continuous.
\[ P \left( (Y_1, \ldots, Y_n) \in C \right) = \int_{(y_1, \ldots, y_n) \in C} f_{Y_1, \ldots, Y_n}(y_1, \ldots, y_n) \, dy_1 \cdots dy_n \]

\( f_{\tilde{Z}}(\tilde{x}) \) is the joint PDF (probability density function) of \( \tilde{Z} \). More compactly,

\[ P \left( \tilde{Z} \in C \right) = \int_{\tilde{z} \in C} f_{\tilde{Z}}(\tilde{z}) \, d\tilde{z} \]

Consequence of this def.

1. If the joint dist. of \( \tilde{Z} \) is continuous, then

\[ f_{\tilde{Z}}(\tilde{x}) = \frac{1}{d\tilde{y}_1 \cdots d\tilde{y}_n} F_{\tilde{Z}}(\tilde{x}) \]

mixed discrete/continuous

random vectors behave just as they do with 2 rv.

more realistically, it would

Example clinical be unknown, and you (trial (continued)) can think about the
joint dist. of \((B_1, \theta) = (B_i, \ldots, B_n, \theta)\), in which the \(B_i\) are discrete and \(0 < \theta < 1\) is continuous.

Marginal distributions

If you know the joint PDF of \(B_i\), you can work out the marginal distribution of any subset of \((Y_1, \ldots, Y_n)\) by integrating \(f_{Y_i}(y_i)\) over the elements of \((Y_1, \ldots, Y_n)\) that are not in the subset.

Example

\[
\begin{align*}
Y &= (Y_1, Y_2, Y_3, Y_4) \\
&= (Y_1, Y_2, Y_3, Y_4) \\
\text{so that} &
\end{align*}
\]

\[
\begin{align*}
f_{Y_1}(y_1) &= \sum \sum \sum f_{Y_i}(y_i) & y_i \neq y_1, y_4, y_3 \\
&= \sum \sum \sum f_{Y_i}(y_i) & y_i \neq y_1, y_4, y_3 \\
&= \sum f_{Y_i}(y_i) & y_1, y_4, y_3
\end{align*}
\]

Similarly, you can work out a marginal CDF by sending the other components...
to \infty: \text{ for example}

\[ F_{\xi_i}(x) = \mathbb{P}(\xi_i \leq x) = \mathbb{P}(\xi_1 \leq x, \xi_2 \leq \infty, \ldots, \xi_n \leq \infty) \]

\[ = \lim_{y_1 \to \infty, \ldots, y_n \to \infty} F_{\xi_i}(x) \quad \text{Definition} \]

In rvs \( \xi_1, \ldots, \xi_n \) are independent if

\( n \) rvs \( \xi_1, \ldots, \xi_n \) are independent if non-identically

for any sets \( A_1, \ldots, A_n \) of real numbers

\[ P(\xi_1 \in A_1, \ldots, \xi_n \in A_n) = \prod_{i=1}^{n} P(\xi_i \in A_i) \]

Immediate consequence:

1. \( \xi_1, \ldots, \xi_n \) independent iff

\[ F_{\xi_1}(x) = \prod_{i=1}^{n} F_{\xi_i}(x) \]

2. \( \xi_1, \ldots, \xi_n \) independent iff

\[ f_{\xi_1}(x) = \prod_{i=1}^{n} f_{\xi_i}(x) \]

\[ (9.\text{m}) \]
Def. Starting with a univariate PDF or PMF $f_{Z_i}$, a set $\{Z_1, ..., Z_n\}$ forms a random sample from $f_{Z_i}$ if the $Z_i$ are independent and all of them have marginal PDF or PMF $f_{Z_i}$, i.e., if the $Z_i$ are an independent identically distributed (IID) sample from $f_{Z_i}$.

Example:

Deer at USC:

- Some have a disease (chronic wasting disease)

Population:

- All deer living within USC boundary

Sample:

- The observed deer

\[ N = 7 \]
\[ \theta = 0.5 \]
\[ \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} \gamma_i \]

\[ \hat{\theta} = \bar{Y} \]

\[ \bar{Y} \] is the estimate of \( \theta \).
Shortcut for the diagram: \[ \text{Definition} \]
\[
\begin{align*}
(\mathbf{X}, \theta) & \sim \text{Bernoulli}(\theta) \\
(z_1, \ldots, z_n) & \\
\mathbf{Z} & = (\mathbf{Z}, \theta) \\
\mathbf{Z}_k & = (z_1, \ldots, z_{n-k})
\end{align*}
\]

Start with random vector \( \mathbf{Z} \); partition it into \( 2 \) subvectors \( \mathbf{Z} = (\mathbf{Z}, \theta) \), \( \mathbf{Z} = (z_1, \ldots, z_k) \), \( \mathbf{Z}_k = (z_1, \ldots, z_{n-k}) \) \( \sim \) \( 1 \leq k \leq n \).

Then for every point \( z \) for which \( f_{\mathbf{Z}_k}(z) > 0 \), the conditional distribution of \( \mathbf{Z} \) given \( \mathbf{Z}_k \) is

\[
\begin{align*}
f_{\mathbf{Z}_k | \mathbf{Z}_k}(z_{1:k}) & = f_{\mathbf{Z}_k}(z_{1:k}) \\
& = f_{\mathbf{Z}_k}(z_{1:k}), \quad z_{1:k} \in \mathbb{R}^k
\end{align*}
\]

From which

\[
f_{\mathbf{Z} | \mathbf{Z}_k}(z_{1:k}, z_{k+1:n}) = f_{\mathbf{Z}_k}(z_{1:k}) f_{\mathbf{Z} | \theta}(z_{k+1:n}).
\]
You'll recall that if $A$ is an event and you're trying to compute $P(A)$ & it's hard, one idea is to find another aspect of the world $\mathcal{B}$ upon which $A$ depends, such that the events $\mathcal{B}_1, \ldots, \mathcal{B}_n$ form a partition; then

$$P(A) = \sum_{i=1}^{n} P(A \cap \mathcal{B}_i) = \sum_{i=1}^{n} P(\mathcal{B}_i)P(A|\mathcal{B}_i)$$

This has an analogue with continuous rvs:

$$f_{\mathcal{B}}(x) = \int \cdots \int_{R^{n-k}} f_{\mathcal{B}_i}(z) f_{X|\mathcal{B}_i}(x|z) \, dz \cdots dz$$

using the notation in the definition of conditional distributions.
Multivariate Bayes's Theorem

The usual application of this in statistics is as follows.

Def. \( \mathbf{Z} \) a random vector with multivariate distribution \( f_{\mathbf{Z}}(\mathbf{z}) \); then random variables \( Z_1, \ldots, Z_n \) are conditionally independent given \( \mathbf{Z} \) if for all \( \mathbf{z} \) with \( f_{\mathbf{Z}}(\mathbf{z}) > 0 \),

\[
\frac{f_{\mathbf{Z}}(\mathbf{z} \mid \mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{z})} = \prod_{i=1}^{n} f_{Z_i}(z_i \mid \mathbf{Z}).
\]
Earlier, we agreed that, if \( \theta \) is unknown to you, the results of the coin tosses \( Z_1, Z_2, \ldots \) are dependent, because there is useful information in any subset of them for predicting any other subset, but \( \mathbb{E} \) the \( Z_i \) become conditionally independent given \( \theta \), because once you know \( \theta \), there's no longer any useful information in the \( Z_i \) to predict other \( Z_j \).
This is why — in both the clinical trial example & the (nuts & bolts) example — we model the data values \( Z \) as
\[
(Z_i | \theta) \sim \text{Bernoulli}(\theta).
\]

Functions of a rv

\( (\text{univariate}) \)

**Case 1:** \( X \) discrete rv with \( P(X = x) \); discrete \( Y = h(X) \) for some function \( h \) defined on \{ possible values of \( X \) \}. Then
\[
f_Z(y) = P(Y = y) = P(h(X) = y)
\]
\[
= \sum_{\{X: h(X) = y\}} f_X(x)
\]

Example

\( X \sim \text{Uniform \{1, 2, ..., 9\}} \)

The median of this distribution is 5; \( Y = |X - 5| \) keeps track of how far \( X \) is from the median.
The CDF $F_Z(y)$ can be worked out as follows: $F_Z(y) = P(Z \leq y) = P(h(X) \leq y)$

$$F_Z(y) = \int_{\{x: h(x) \leq y\}} f_X(x) \, dx$$

and if $Z$ is also continuous

$$f_Z(y) = \frac{d}{dy} F_Z(y) \quad (\text{at every point where } F_Z \text{ is differentiable})$$
Example: \( \bar{X} \) is rate at which customers arrive in a queue at the bank

Natural to model \( \bar{X} \) as continuous, (also, \( \bar{X} > 0 \)), with CDF \( F_{\bar{X}} \).

Turns out that the average waiting time is \( \bar{X} = \frac{1}{h(x)} \).

You can get the PDF of \( \bar{X} \) in 2 steps:

1. work out CDF of \( \bar{X} \)
2. differentiate with respect to \( \gamma \)

\( \text{(for } \gamma > 0) \)

\[
F_{\bar{X}}(\gamma) = P(\bar{X} \leq \gamma) = P\left( h(x) \leq \gamma \right)
\]

\[= P\left( \frac{1}{\bar{X}} \leq \gamma \right) = P\left( \bar{X} \geq \frac{1}{\gamma} \right) \quad \text{since } \bar{X} \text{ is continuous}
\]

\[= 1 - P\left( \bar{X} < \frac{1}{\gamma} \right) = 1 - P\left( \bar{X} \leq \frac{1}{\gamma} \right)
\]

\[= 1 - F_{\bar{X}}\left( \frac{1}{\gamma} \right) \quad \text{and now}
\]
Example

\( X \sim \text{Uniform } [-1, 1] \) (continuous)

\( Y = X^2 \)

Find PDF of \( Y \)

First note that \( Y \)'s possible values are \([-1, 1]\).

For \(-1 < y < 1\):

1. \( F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x) \, dx \)

Thus

\[
\frac{d}{dy} F_Y(y) = \frac{d}{dy} \left( F_X(\sqrt{y}) - F_X(-\sqrt{y}) \right) = \frac{1}{2 \sqrt{y}}
\]

\[
f_Y(y) = \frac{1}{2 \sqrt{y}}
\]

\( f_X(x) = \begin{cases} \frac{1}{2}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \)