Week 7: Transformations of a Random Variable and Joint Distributions

MATH697

Sahir Bhatnagar

October 17, 2017

McGill University

One Dimensional Change of Variable

Motivation

- Often we need to deal with a transformation $Y = g(X)$ of the random variable X

Motivation

- Often we need to deal with a transformation $Y = g(X)$ of the random variable X
- Here $g(X)$ could be a simple change of time scale.

Motivation

- Often we need to deal with a transformation $Y = g(X)$ of the random variable X
- Here $g(X)$ could be a simple change of time scale.
- If X is in hours and Y is in minutes, then $Y = 60X$

Motivation

- Often we need to deal with a transformation $Y = g(X)$ of the random variable X
- Here $g(X)$ could be a simple change of time scale.
- If X is in hours and Y is in minutes, then $Y = 60X$

Motivation

- Often we need to deal with a transformation $Y = g(X)$ of the random variable X
- Here $g(X)$ could be a simple change of time scale.
- If X is in hours and Y is in minutes, then $Y = 60X$

Can we get the pdf of Y from the pdf of X ?

Discrete Change of Variable

Discrete Case

- Let X is a *discrete* RV

Discrete Case

- Let X is a *discrete* RV
- We want to compute the probability that $Y = y$.

Discrete Case

- Let X is a *discrete* RV
- We want to compute the probability that $Y = y$.
- To do so, we need to compute the probability of the set consisting of all the x values satisfying $g(x) = y$, given by

$$P(X \in \{x : g(x) = y\})$$

Discrete Case

- Let X is a *discrete* RV
- We want to compute the probability that $Y = y$.
- To do so, we need to compute the probability of the set consisting of all the x values satisfying $g(x) = y$, given by

$$P(X \in \{x : g(x) = y\})$$

- Let X is a *discrete* RV
- We want to compute the probability that $Y = y$.
- To do so, we need to compute the probability of the set consisting of all the x values satisfying $g(x) = y$, given by

$$P(X \in \{x : g(x) = y\})$$

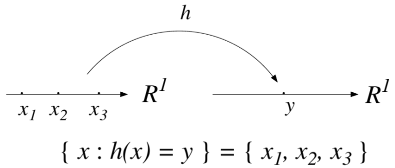


Figure 2.6.1: An example where the set of x values that satisfy $h(x) = y$ consists of three points x_1, x_2, and x_3.

Example 1 (Flipping Coins)

Let X be the number of heads when flipping three fair coins. Let

$$Y = \begin{cases} 1 & X \geq 1 \\ 0 & X = 0 \end{cases}$$

Find the PDF of Y

Solution: on board

Theorem for Discrete Change of Variable

Theorem 2 (Discrete Change of Variable)

Let X be a discrete RV with PMF $f_X(x)$. Let $Y = h(X)$, where $g : \mathbb{R} \to \mathbb{R}$ is some function. Then Y is also discrete and its PMF $f_Y(y)$ satisfies

$$f_Y(y) = \sum_{x \in g^{-1}\{y\}} f_X(x),$$

where $g^{-1}\{y\}$ is the set of all real numbers x with $g(x) = y$

Discrete Change of Variable Example 2

Example 3 (Fair six-sided die)

Let X be the number showing on a fair six-sided die, so that $P(X = x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$. Let

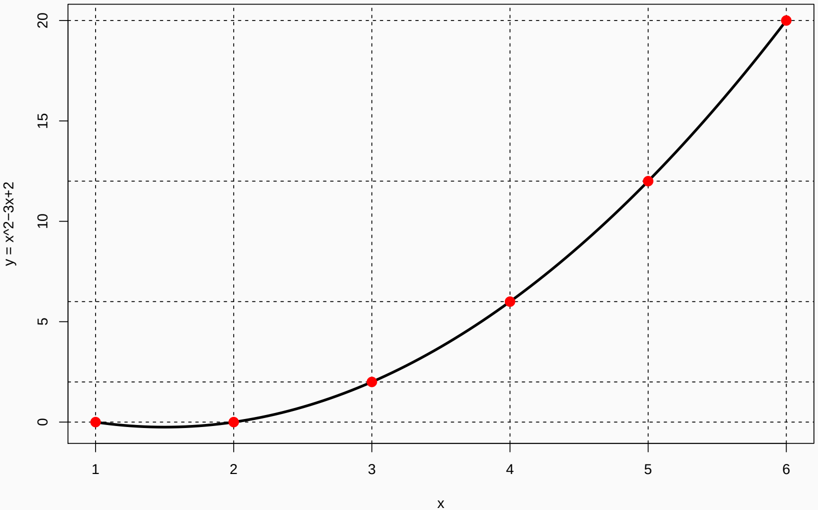
$$Y = X^2 - 3X + 2$$

Find the PDF of Y

Solution: on board

Discrete Change of Variable Example 2

```r
curve(x^2-3*x+2, from = 1, to = 6, ylab = "y = x^2-3x+2", lwd = 3)
abline(h=(c(0,2,6,12,20)), lty = 2); abline(v=1:6, lty = 2)
points(1:6, c(0,0,2,6,12,20), pch = 19, col = "red", cex = 1.5)
```



Example 4 (Binomial Distribution)

Let $X \sim Binomial(n, p)$, and consider the RV $Y = n - X$, which corresponds to the number of failures. Show that $Y \sim Binomial(n, 1 - p)$

Solution: on board

Continuous Change of Variable

Continous Change of Variable

If X is continuous and $Y = g(X)$, then the situation is more complicated as Y might not be continuous at all as the following example shows

Example 5 (Discrete Transformation of a Continuous Variable)

Let $X \sim Uniform(0, 1)$. Let $Y = g(X)$, where

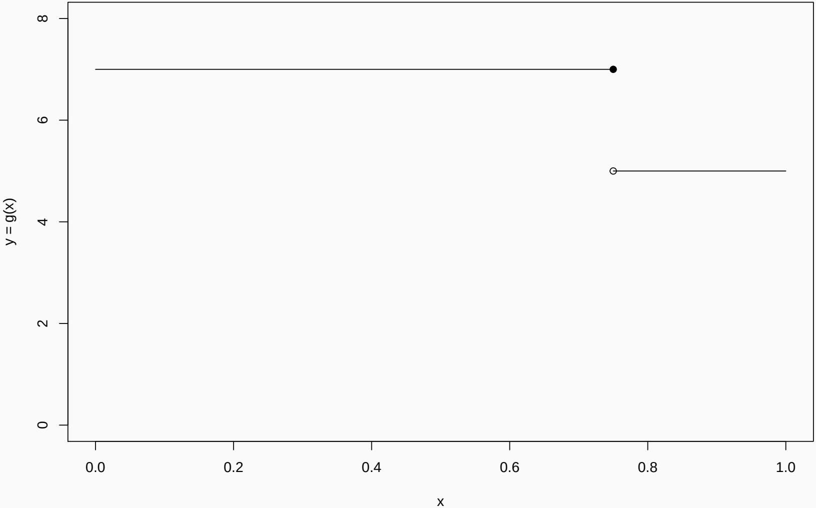
$$g(x) = \begin{cases} 7 & x \leq 3/4 \\ 5 & x > 3/4 \end{cases}$$

Hence, Y is discrete with PMF

$$f_Y(y) = \begin{cases} 3/4 & y = 7 \\ 1/4 & y = 5 \\ 0 & y \neq 5, 7 \end{cases}$$

Example 1

```
plot(0:1,type = "n", xlim = 0:1, ylim = c(0,8), ylab = "y = g(x)", xlab = "x")
segments(x0 = 0, y0 = 7, x1 = 3/4); segments(x0 = 3/4, y0 = 5, x1 = 1)
points(3/4,7, pch=19);points(3/4,5)
```



Example 6 (Interval between calls to a 911 center)

The interval X in minutes between calls to a 911 center is exponentially distributed with mean 2 min. What is the PDF of the number of seconds?

Solution: on board

Example 7 (Transformation of a Uniform RV)

Let $X \sim \mathit{Uniform}(0, 1)$ and let $Y = 3X$. What is the distribution of Y?

Solution: on board

Theorem 8 (Monotone Transformations)

Let X be an absolutely continuous RV, with density function f_X. Let $Y = g(X)$, where $g : \mathbb{R} \to \mathbb{R}$ is a differentiable (smooth) that is strictly ***increasing*** *or* ***decreasing****, i.e,* ***monotonic****, so it has an inverse function g^{-1}. Then Y is also continuous with PDF*

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{\partial}{\partial y} g^{-1}(y) \right| \tag{1}$$

Proof: on board

Example 9 (Transformation of a Uniform RV)

Let $X \sim Uniform(0,1)$ and let $Y = 3X$. Then let $Y = g(X) = 3X$.

Note that g is strictly increasing because if $x < y$ then $g(x) < g(y)$.

Hence we may apply the theorem.

$g^{-1}(y) = y/3$ and $\frac{\partial}{\partial y} g^{-1}(y) = 1/3$. Applying the formula given by (1) we get

$$f_Y(y) = f_X\left(y/3\right) \cdot \frac{1}{3} = \begin{cases} 1/3 & 0 \leq y \leq 3 \\ 0 & otherwise \end{cases}$$

Example 10 (Transformation of a Normal RV)

Let $X \sim N(0, 1)$ and let $Y = 2X + 5$. What is the distribution of Y?

Solution: on board

Example 11 (Instance of when the Theorem can't be used)

Let $f_X(x) = \frac{x+1}{8}$, for $-1 < x < 3$ and $Y = X^2$. The transformation is not monotonic. Why? What is the PDF of Y?

Solution: on board

Joint Probability Distributions

- When we first introduced RVs and their distributions, we noted that the individual distributions of two RVs do not tell us anything about whether the RVs are independent or dependent.

- When we first introduced RVs and their distributions, we noted that the individual distributions of two RVs do not tell us anything about whether the RVs are independent or dependent.
- For example, two $Bernoulli(1/2)$ RVs X and Y could be independent if they indicate Heads on two different coin flips, or dependent if they indicate Heads and Tails respectively on the same coin flip.

Example 12 (Bernoulli Flips)

Let $X \sim Bernoulli(1/2)$ so that $P(X = 0) = P(X = 1) = 1/2$. Let $Y_1 = X$ and $Y_2 = 1 - X$. Then we clearly have $Y_1 \sim Bernoulli(1/2)$ and $Y_2 \sim Bernoulli(1/2)$. On the other hand, the **relationship** between X and Y_1 is very different from the relationship between X and Y_2. For example, if we know that $X = 1$, then we also know that $Y_1 = 1$ but $Y_2 = 0$.

Hence, merely knowing that X, Y_1, and Y_2 all have the distribution $Bernoulli(1/2)$ does not give us complete information about the relationships among these random variables

- Thus, although the PMF of X is a complete blueprint for X and the PMF of Y is a complete blueprint for Y, these individual PMFs are missing important information about how the two RVs are related.

Introduction

Of course, in real life, we usually care about the relationship between multiple RVs in the same experiment. To give just a few examples:

1. **Medicine**: To evaluate the effectiveness of a treatment, we may take multiple measurements per patient; an ensemble of blood pressure, heart rate, and cholesterol readings can be more informative than any of these measurements considered separately.

Introduction

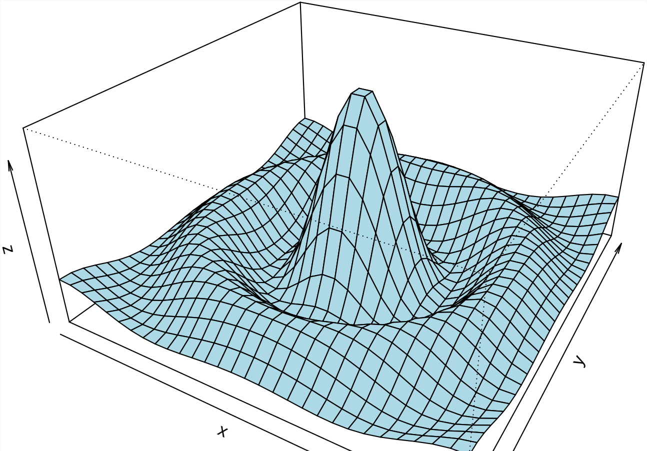
Of course, in real life, we usually care about the relationship between multiple RVs in the same experiment. To give just a few examples:

1. **Medicine**: To evaluate the effectiveness of a treatment, we may take multiple measurements per patient; an ensemble of blood pressure, heart rate, and cholesterol readings can be more informative than any of these measurements considered separately.
2. **Genetics**: To study the relationships between various genetic markers and a particular disease, if we only looked separately at distributions for each genetic marker, we could fail to learn about whether an interaction between markers is related to the disease.

3. **Time series**: To study how something evolves over time, we can often make a series of measurements over time, and then study the series jointly. There are many applications of such series, such as global temperatures, stock prices, or national unemployment rates. The series of measurements considered jointly can help us deduce trends for the purpose of forecasting future measurements.

Example of a Multivariate Distribution

```
x <- seq(-10, 10, length= 30); y <- x
f <- function(x, y) { r <- sqrt(x^2+y^2); 10 * sin(r)/r }
z <- outer(x, y, f); z[is.na(z)] <- 1
op <- par(bg = "white")
persp(x, y, z, theta = 30, phi = 30, expand = 0.5, col = "lightblue")
```



Joint, Marginal and Conditional

3 Key Concepts

- The three key concepts for this section are joint, marginal, and conditional distributions.

3 Key Concepts

- The three key concepts for this section are joint, marginal, and conditional distributions.
- Recall that the distribution of a single RV X provides complete information about the probability of X falling into any subset of the real line.

- Analogously, the joint distribution of two RVs X and Y provides complete information about the probability of the vector (X, Y) falling into any subset of the plane.

- Analogously, the joint distribution of two RVs X and Y provides complete information about the probability of the vector (X, Y) falling into any subset of the plane.
- The marginal distribution of X is the individual distribution of X, ignoring the value of Y

- Analogously, the joint distribution of two RVs X and Y provides complete information about the probability of the vector (X, Y) falling into any subset of the plane.
- The marginal distribution of X is the individual distribution of X, ignoring the value of Y
- The conditional distribution of X given $Y = y$ is the updated distribution for X after observing $Y = y$.

- Analogously, the joint distribution of two RVs X and Y provides complete information about the probability of the vector (X, Y) falling into any subset of the plane.
- The marginal distribution of X is the individual distribution of X, ignoring the value of Y
- The conditional distribution of X given $Y = y$ is the updated distribution for X after observing $Y = y$.
- We'll look at these concepts in the discrete case first, then extend them to the continuous case.

Discrete Distributions - Joint CDF

Definition 13 (Joint CDF)

The joint CDF of RVs X and Y is the function $F_{X,Y}$ given by

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

The joint CDF of n RVs is defined analogously.

Discrete Distributions - Joint CDF

Definition 13 (Joint CDF)

The joint CDF of RVs X and Y is the function $F_{X,Y}$ given by

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

The joint CDF of n RVs is defined analogously.

Unfortunately, the joint CDF of discrete RVs is not a well-behaved function; as in the univariate case, it consists of jumps and flat regions. For this reason, with discrete RVs we usually work with the joint PMF, which also determines the joint distribution and is much easier to visualize.

Definition 14 (Joint PMF)

The joint PMF of RVs X and Y is the function $f_{X,Y}$ given by

$$f_{X,Y}(x, y) = P(X = x, Y = y).$$

The joint CDF of n RVs is defined analogously.

Just as univariate PMFs must be nonnegative and sum to 1, we require valid joint PMFs to be nonnegative and sum to 1, where the sum is taken over all possible values of X and Y:

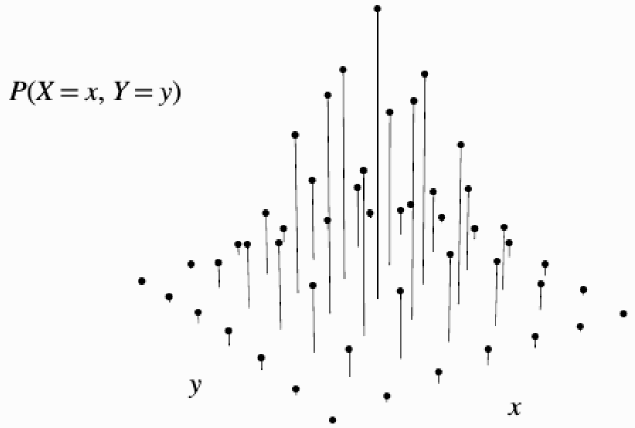
$$\sum_x \sum_y P(X = x, Y = y) = 1$$

The joint PMF determines the distribution because we can use it to find the probability of the event $(X, Y) \in A$ for any set A in the plane. All we have to do is sum the joint PMF over A:

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

Discrete Distributions - Joint PMF

The figure shows a sketch of what the joint PMF of two discrete RVs could look like. The height of a vertical bar at (x, y) represents the probability $P(X = x, Y = y)$. For the joint PMF to be valid, the total height of the vertical bars must be 1.



Definition 15 (Marginal PMF)

For discrete RVs X and Y, the marginal PMF of X is

$$f_X(x) = P(X = x) = \sum_y P(X = x, Y = y)$$

The marginal PMF of X is the PMF of X, viewing X individually rather than jointly with Y.

Definition 15 (Marginal PMF)

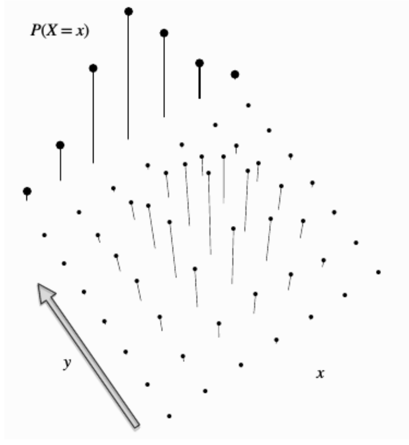
For discrete RVs X and Y, the marginal PMF of X is

$$f_X(x) = P(X = x) = \sum_y P(X = x, Y = y)$$

The marginal PMF of X is the PMF of X, viewing X individually rather than jointly with Y.

The above equation follows from the axioms of probability (we are summing over disjoint cases). The operation of summing over the possible values of Y in order to convert the joint PMF into the marginal PMF of X is known as marginalizing out Y.

Each column of the joint PMF corresponds to a fixed x and each row corresponds to a fixed y. For any x, the probability $P(X = x)$ is the total height of the bars in the corresponding column of the joint PMF: we can imagine taking all the bars in that column and stacking them on top of each other to get the marginal probability. Repeating this for all x, we arrive at the marginal PMF, depicted in bold.



Discrete Distributions - Marginal PMF Remarks

- Similarly, the marginal PMF of Y is obtained by summing over all possible values of X. So given the joint PMF, we can marginalize out Y to get the PMF of X, or marginalize out X to get the PMF of Y.

Discrete Distributions - Marginal PMF Remarks

- Similarly, the marginal PMF of Y is obtained by summing over all possible values of X. So given the joint PMF, we can marginalize out Y to get the PMF of X, or marginalize out X to get the PMF of Y.
- But if we only know the marginal PMFs of X and Y, there is no way to recover the joint PMF without further assumptions.

- Similarly, the marginal PMF of Y is obtained by summing over all possible values of X. So given the joint PMF, we can marginalize out Y to get the PMF of X, or marginalize out X to get the PMF of Y.
- But if we only know the marginal PMFs of X and Y, there is no way to recover the joint PMF without further assumptions.
- It is clear how to stack the bars in the previous figure, but very unclear how to unstack the bars after they have been stacked

Discrete Distributions - Marginal PMF Remarks

- Another way to go from joint to marginal distributions is via the joint CDF. In that case, we take a limit rather than a sum: the marginal CDF of X is

$$F_X(x) = P(X \leq x) = \lim_{y \to \infty} P(X \leq x, Y \leq y) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

  However, as mentioned above it is usually easier to work with joint PMFs.

Definition 16 (Conditional PMF)

For discrete RVs X and Y, the conditional PMF of Y given $X = x$ is

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

This is viewed as a function of y for fixed x.

Definition 16 (Conditional PMF)

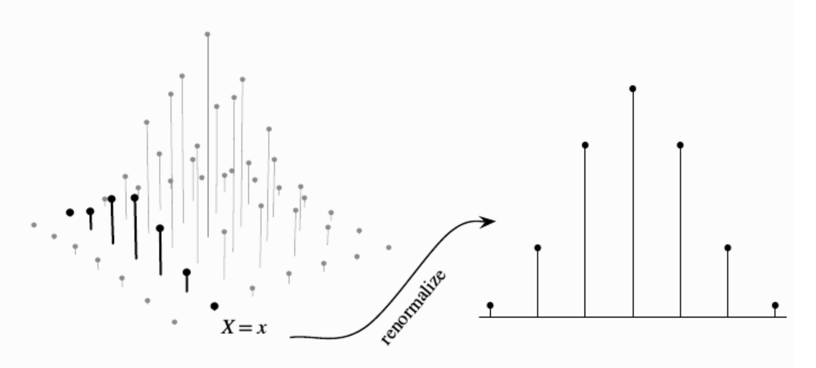
For discrete RVs X and Y, the conditional PMF of Y given $X = x$ is

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

This is viewed as a function of y for fixed x.

Note that the conditional PMF (for fixed x) is a valid PMF. So we can define the conditional expectation of Y given $X = x$, denoted by $E(Y \mid X = x)$, in the same way that we defined $E(Y)$ except that we replace the PMF of Y with the conditional PMF of Y.

Conditional PMF of Y given $X = x$. The conditional PMF $P(Y = y \mid X = x)$ is obtained by renormalizing the column of the joint PMF that is compatible with the event $X = x$.



We can also relate the conditional distribution of Y given $X = x$ to that of X given $Y = y$, using Bayes' rule:

$$P\big(Y = y \mid X = x\big) = \frac{P\big(X = x \mid Y = y\big)P\big(Y = y\big)}{P\big(X = x\big)}$$

We can also relate the conditional distribution of Y given $X = x$ to that of X given $Y = y$, using Bayes' rule:

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

And using the law of total probability, we have another way of getting the marginal PMF: the marginal PMF of X is a weighted average of conditional PMFs $P(X = x \mid Y = y)$, where the weights are the probabilities $P(Y = y)$:

$$P(X = x) = \sum_y P(X = x \mid Y = y)P(Y = y)$$

Example

Example 17 (2 x 2 table)

The simplest example of a discrete joint distribution is the case where X and Y are both Bernoulli RVs. In this case, the joint PMF is fully specified by the four values $P(X = 1, Y = 1)$, $P(X = 0, Y = 1)$, $P(X = 1, Y = 0)$, and $P(X = 0, Y = 0)$, so we can represent the joint PMF of X and Y using a 2×2 table.

Example

Example 17 (2 x 2 table)

The simplest example of a discrete joint distribution is the case where X and Y are both Bernoulli RVs. In this case, the joint PMF is fully specified by the four values $P(X = 1, Y = 1)$, $P(X = 0, Y = 1)$, $P(X = 1, Y = 0)$, and $P(X = 0, Y = 0)$, so we can represent the joint PMF of X and Y using a 2×2 table.

This very simple scenario actually has an important place in statistics, as these so-called contingency tables are often used to study whether a treatment is associated with a particular outcome. In such scenarios, X may be the indicator of receiving the treatment, and Y may be the indicator of the outcome of interest.

Example

Example 18 (2 x 2 table)

For example, suppose we randomly sample an adult male from the United States population. Let X be the indicator of the sampled individual being a current smoker, and let Y be the indicator of his developing lung cancer at some point in his life. Then the following table could represent the joint PMF of X and Y.

	Y = 1	Y = 0
X = 1	5/100	20/100
X = 0	3/100	72/100

Verify that this is a true joint distribution. Compute the marginal probabilities, then the marginal distributions. Suppose now we observe $X = 1$. What is their risk for lung cancer?

Independence of discrete RVs

Definition 19 (Independence of discrete RVs)

Random variables X and Y are independent if for all x and y,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

. If X and Y are discrete, this is equivalent to the condition

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all x and y, and it is also equivalent to the condition

$$P(Y = y \mid X = x) = P(Y = y)$$

for all y and all x such that $P(X = x) > 0$

Independence of discrete RVs

- Using the terminology from this chapter, the definition says that for independent RVs, the joint CDF factors into the product of the marginal CDFs, or that the joint PMF factors into the product of the marginal PMFs.

Independence of discrete RVs

- Using the terminology from this chapter, the definition says that for independent RVs, the joint CDF factors into the product of the marginal CDFs, or that the joint PMF factors into the product of the marginal PMFs.
- Remember that in general, the marginal distributions do not determine the joint distribution: this is the entire reason why we wanted to study joint distributions in the first place!

Independence of discrete RVs

- Using the terminology from this chapter, the definition says that for independent RVs, the joint CDF factors into the product of the marginal CDFs, or that the joint PMF factors into the product of the marginal PMFs.
- Remember that in general, the marginal distributions do not determine the joint distribution: this is the entire reason why we wanted to study joint distributions in the first place!
- But in the special case of independence, the marginal distributions are all we need in order to specify the joint distribution; we can get the joint PMF by multiplying the marginal PMFs.

Independence of discrete RVs

- Another way of looking at independence is that all the conditional PMFs are the same as the marginal PMF. In other words, starting with the marginal PMF of Y, no updating is necessary when we condition on $X = x$, regardless of what x is.

Independence of discrete RVs

- Another way of looking at independence is that all the conditional PMFs are the same as the marginal PMF. In other words, starting with the marginal PMF of Y, no updating is necessary when we condition on $X = x$, regardless of what x is.
- There is no event purely involving X that influences our distribution of Y , and vice versa.

Example Revisited

Example 20 (2 x 2 table)

Are X and Y independent?

	Y = 1	Y = 0
X = 1	5/100	20/100
X = 0	3/100	72/100

Explain both ways in which this can be shown.