Chapter 1.1

Probability

Review of probability

- The crux of Bayesian statistics is to compute the posterior distribution, i.e., the uncertainty distribution of the parameters (θ) after observing the data ($\mathbf{Y}$)

- This is the conditional distribution of θ given $\mathbf{Y}$

- Therefore, we need to review the probability concepts that lead to the conditional distribution of one variable conditioned on another
  1. Probability mass (PMF) and density (PDF) functions
  2. Joint distributions
  3. Marginal and conditional distributions

Random variables

- X (capital) is a random variable
- We want to compute the probability that X takes on a specific value x (lowercase)
- This is denoted $\text{Prob}(X = x)$
- We also might want to compute the probability of X being in a set $\mathcal{A}$
- This is denoted $\text{Prob}(X \in \mathcal{A})$
- The set of possible value that X can take on is called its support, $\mathcal{S}$

Random variables - example

Example 1: X is the roll of a die

- The support is $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$
- $\text{Prob}(X = 1) = 1/6$

Example 2: X is a newborn baby's weight

- The support is $\mathcal{S} = (0, \infty)$
- $\text{Prob}(X \in [0, \infty]) = 1$

What is probability?

Objective (associated with frequentist)

- Prob($X = x$) as a purely mathematical statement
- If we repeatedly sampled X, the the proportion of draws equal to x converges to Prob($X = x$)

Subjective (associated with Bayesian)

- Prob($X = x$) represents an individual's degree of belief
- Often quantified as the amount an individual would be willing to wager that X will be x

A Bayesian analysis makes use of both of these concepts

What is uncertainty?

Aleatoric uncertainty (likelihood)

- Uncontrollable randomness in the experiment
- For example, the results of a fair coin flip can never be predicted with certainty

Epistemic uncertainty (prior/posterior)

- Uncertainty about a quantity that could theoretically be known
- For example, if we flipped a coin infinitely-many times we could know the true probability of a head

A Bayesian analysis makes use of both of these concepts

Probability versus statistics

Probability is the forward problem

- We assume we know how the data are being generated and compute the probability of events
- For example, what is the probability of flipping 5 straight heads if the coin is fair?

Statistics is the inverse problem

- We use data to learn about the data-generating mechanism
- For example, if we flipped five straight head, can we conclude the coin is biased?

Any statistical analysis obviously relies on probability

Univariate distributions

- We often distinguish between **discrete** and **continuous** random variables

- The random variable X is discrete if its support $\mathcal{S}$ is countable

- Examples:

  $X \in \{0, 1, 2, 3\}$ is the number of successes in 3 trials

  $X \in \{0, 1, 2, ...\}$ is the number users that visit a website

Univariate distributions

- We often distinguish between **discrete** and **continuous** random variables
- The random variable X is continuous if its support $\mathcal{S}$ is uncountable
- Examples with $\mathcal{S} = (0, \infty)$:

  $X > 0$ is weight of a baby

  $X > 0$ is the wind speed

Discrete univariate distributions

- If X is discrete we describe its distribution with its **probability mass function** (PMF)
- The PMF is $f(x) = \text{Prob}(X = x)$
- The domain of X is the set of x with $f(x) > 0$
- We must have $f(x) \geq 0$ and $\sum_x f(x) = 1$
- The mean is $\text{E}(X) = \sum_x xf(x)$
- The variance is $\text{V}(X) = \sum_x [x - E(X)]^2 f(x)$
- The last three sums are over X's domain

Discrete univariate distributions

- If X is discrete we describe its distribution with its **probability mass function** (PMF)
- The PMF is $f(x) = \text{Prob}(X = x)$
- The support of X, $\mathcal{S}$, is the set of x with $f(x) > 0$
- For f to be a valid PMF we must have $f(x) \geq 0$ and $\sum_x f(x) = 1$

Discrete univariate distributions

- Distributions can be summarized by their mean and variance
- The mean is $\mathrm{E}(X) = \sum_x xf(x)$
- The variance is $\mathrm{V}(X) = \sum_x [x - E(X)]^2 f(x)$
- The sums are over X's domain
- These summaries of the distribution can be estimated by the sample mean ($\bar{X}$) and variance (s^2)

Parametric families of distributions

- A statistical analysis typically proceeds by selecting a PMF that seems to match the distribution of a sample
- We rarely know the PMF exactly, but we assume it is from a parametric family of distributions
- For example, Binomial(10,0.5) and Binomial(4,0.1) are different but both from the binomial family
- A family of distributions have the same equation for the PMF but differ by some unknown parameters θ
- We must estimate these parameters

Example: $X \sim$ Bernoulli(θ)

- Example: X is a success (1) or failure (0)
- Domain: $X \in \{0, 1\}$ (i.e., X is binary)
- PMF: $P(X = 0) = 1 - \theta$ and $P(X = 1) = \theta$
- Parameter: $\theta \in [0, 1]$ is the success probability
- Mean: $E(X) = \sum_x xf(x) = 0(1 - \theta) + 1\theta = \theta$
- Variance:

$$V(X) = \sum_x (x-\theta)^2 f(x) = (0-\theta)^2(1-\theta)+(1-\theta)^2\theta = \theta(1-\theta)$$

Example: $X \sim$ Binomial(N, θ)

- Example: X is a number of successes in N trials
- Domain: $X \in \{0, 1, ..., N\}$
- PMF: $P(X = x) = \binom{N}{x}\theta^x(1-\theta)^{N-x}$
- Parameter: $\theta \in [0, 1]$ is the success probability of each trial
- Mean: $\mathrm{E}(X) = \sum_{x=0}^{N} xf(x) = n\theta$
- Variance: $\mathrm{V}(X) = n\theta(1-\theta)$

Example: $X \sim \text{Poisson}(N\theta)$

- Example: X is the number events that occur in N units of time
- Often the distribution is presented with $N = 1$
- Domain: $X \in \{0, 1, 2, ...\}$
- PMF: $P(X = x) = \frac{\exp(-N\theta)(N\theta)^x}{x!}$
- Parameter: θ is the expected number of events per unit of time
- Mean: $E(X) = N\theta$
- Variance: $V(X) = N\theta$

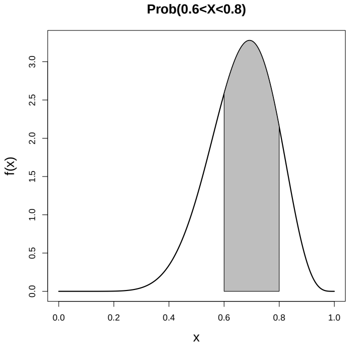
Continuous univariate distributions

- If X is continuous we describe its distribution with the probability density function (PDF) $f(x) \geq 0$
- Since there are uncountably many possible values, the probability of any one value must be zero
- In lab this week we will construct an informal proof of this
- Therefore, $P(X = x) = 0$ for all x, and so the PMF is meaningless

Continuous univariate distributions

- Probabilities are computed as areas under the PDF curve

$$\text{Prob}(l < X < u) = \int_l^u f(x)dx$$



- Therefore, to be valid $f(x)$ must satisfy $f(x) \geq 0$ and

$$\text{Prob}(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

Continuous univariate distributions

- The domain is the set of x values with $f(x) > 0$
- The mean and the variance are defined similarly to the discrete case but with the sums replaced by integrals
- The mean is

$$\mathrm{E}(X) = \int x f(x) dx$$

- The variance is

$$\mathrm{V}(X) = \int [x - E(X)]^2 f(x) dx$$

Example: $X \sim$ Normal(μ, σ^2)

- Example: X is an IQ score
- Domain: $X \in (-\infty, \infty)$
- PDF: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$
- Parameters: μ is the mean, $\sigma^2 > 0$ is the variance
- Mean: $E(X) = \mu$
- Variance: $V(X) = \sigma^2$

Example: $X \sim \text{Gamma}(a, b)$

- Example: X is a height
- Domain: $X \in (0, \infty)$
- PDF: $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$
- Parameters: $a > 0$ is the shape, $b > 0$ is the rate
- Mean: $\text{E}(X) = \frac{a}{b}$
- Variance: $\text{V}(X) = \frac{a}{b^2}$
- Be careful: Sometimes the PDF is given as

$$f(x) = \frac{1}{\Gamma(a) b^a} x^{a-1} \exp(-x/b)$$

Example: $X \sim \text{InverseGamma}(a, b)$

- If $Y \sim \text{Gamma}(a, b)$ and $X = 1/Y$, then $X \sim \text{InverseGamma}(a, b)$
- Domain: $X \in (0, \infty)$
- PDF: $f(x) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp(-b/x)$
- Parameters: $a > 0$ is the shape, $b > 0$ is the rate
- Mean: $\text{E}(X) = \frac{b}{a-1}$ if $a > 1$
- Variance: $\text{V}(X) = \frac{b^2}{(a-1)^2(a-2)}$ if $a > 2$
- Be careful: Sometimes the PDF is given as

$$f(x) = \propto x^{-a-1} \exp[-1/(bx)]$$

Example: $X \sim \text{Beta}(a, b)$

- Example: X is a probability
- Domain: $X \in [0, 1]$
- PDF: $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$
- Parameters: $a > 0$ and $b > 0$
- Mean: $\text{E}(X) = \frac{a}{a+b}$
- Variance: $\text{V}(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Joint distributions

- $\mathbf{X} = (X_1, ..., X_p)$ is a random vector (vectors and matrices should be in bold).

- For notational convenience, let's consider only $p = 2$ random variables X and Y.

- (X, Y) is discrete if it can take on a countable number of values, such as

  X = number of hearts and Y = number of face cards.

- (X, Y) is continuous if it can take on an uncountable number of values, such as

  X = birthweight and Y = gestational age.

Discrete random variables

- The joint PMF is

$$f(x, y) = \text{Prob}(X = x, Y = y)$$

- Example: patients select a dose and are followed to determine whether they develop a tumor

- $X \in \{5, 10, 20\}$ is the dose; $Y \in \{0, 1\}$ is 1 if a tumor develops and 0 otherwise

- The joint PMF is

	X		
Y	5	10	20
0	0.469	0.124	0.049
1	0.231	0.076	0.051

Discrete random variables

- The **marginal PMF** for X is

$$f_X(x) = \text{Prob}(X = x) = \sum_y f(x, y)$$

- The **marginal PMF** for Y is

$$f_Y(y) = \text{Prob}(Y = y) = \sum_x f(x, y)$$

- The marginal distribution is the same as univariate distribution as if we ignored the other variable

Discrete random variables

- Example: X = dose and Y = tumor status and

Y	X = 5	10	20
0	0.469	0.124	0.049
1	0.231	0.076	0.051

- Find the marginal PMFs of X and Y

Discrete random variables

- The marginal PMF of X is the column sums

	X			
Y	5	10	20	$f_Y(y)$
0	0.469	0.124	0.049	0.642
1	0.231	0.076	0.051	0.357
$f_X(x)$	0.700	0.200	0.100	

- Marginal distribution of X

$$\begin{aligned} f_X(5) &= \sum_y f(5, y) \\ &= f(5,0) + f(5,1) \\ &= 0.469 + 0.231 = 0.700 \end{aligned}$$

Discrete random variables

- The marginal PMF of X is the row sums

	X			
Y	5	10	20	$f_Y(y)$
0	0.469	0.124	0.049	0.642
1	0.231	0.076	0.051	0.357
$f_X(x)$	0.700	0.200	0.100	

- Marginal distribution of Y

$$\begin{aligned} f_Y(1) &= \sum_x f(x,1) \\ &= f(5,1) + f(10,1) + f(15,1) \\ &= 0.231 + 0.076 + 0.051 = 0.357 \end{aligned}$$

Discrete random variables

- The **Conditional PMF** of Y given X is

$$f(y|x) = \text{Prob}(Y = y|X = x) = \frac{\text{Prob}(X = x, Y = y)}{\text{Prob}(X = x)} = \frac{f(x, y)}{f_X(x)}.$$

- Here x is treated as a fixed number, and so $f(y, x)$ is only a function of y.
- However, we can't use $f(x, y)$ as the PMF for Y because

$$\sum_y f(x, y) = f_X(x) \neq 1$$

- Dividing by $f_X(x)$ makes $f(y|x)$ valid

$$\sum_y f(y|x) = \sum_y \frac{f(y, x)}{f_X(x)} = \frac{\sum_y f(y, x)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$$

Discrete random variables

- Example: X = dose and Y = tumor status
- Find $f(y|X = 5)$

	X			
Y	5	10	20	$f_Y(y)$
0	0.469	0.124	0.049	0.642
1	0.231	0.076	0.051	0.357
$f_X(x)$	0.700	0.200	0.100	

Discrete random variables

- Example: X = dose and Y = tumor status
- Find $f(y|X=5)$

	X			
Y	**5**	10	20	$f_Y(y)$
0	**0.469**	0.124	0.049	0.642
1	**0.231**	0.076	0.051	0.357
$f_X(x)$	**0.700**	0.200	0.100	

- Prob$(Y=0|X=5)$ is
  $f(0|5) = f(5,0)/f_X(5) = 0.469/0.700 = 0.670$
- Prob$(Y=1|X=5)$ is
  $f(1|5) = f(5,1)/f_X(5) = 0.231/0.700 = 0.333$

Discrete random variables

- X and Y are **independent** if

$$f(x, y) = f_X(x) f_Y(y)$$

for all x and y

- Variables are dependent if they are not independent

- Equivalently, X and Y are **independent** if

$$f(x|y) = f_X(x)$$

for all x and y

Discrete random variables

- Notation: $X_1, ..., X_n \overset{iid}{\sim} f(x)$ means that $X_1, ..., X_n$ are independent and identically distributed

- This implies the joint PMF is

$$\text{Prob}(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} f(x_i)$$

- The same notation and definitions of independence apply to continuous random variables

- In this class, assume independence unless otherwise noted

Hurricane proportions by landfall (X) and category (Y)

	Category					
	1	2	3	4	5	Total
US	0.0972	0.0903	0.0694	0.0069	0.0069	0.2708
Not US	0.3194	0.1319	0.1389	0.1181	0.0208	0.7292
Total	0.4167	0.2222	0.2083	0.1250	0.0278	1.0000

Problem: Prove X and Y are dependent

Hurricane proportions by landfall (X) and category (Y)

	Category					
	1	2	3	4	5	Total
US	0.0972	0.0903	0.0694	0.0069	0.0069	0.2708
Not US	0.3194	0.1319	0.1389	0.1181	0.0208	0.7292
Total	0.4167	0.2222	0.2083	0.1250	0.0278	1.0000

Answer: All we need to do is find one (x, y) for which $f(x, y) \neq f_X(x) f_Y(y)$. Let's try $f(US, 1) = 0.972$ versus $f(US) * f(1) = 0.2708 * 0.4167 = 0.1128$. Therefore, X and Y are not independent.

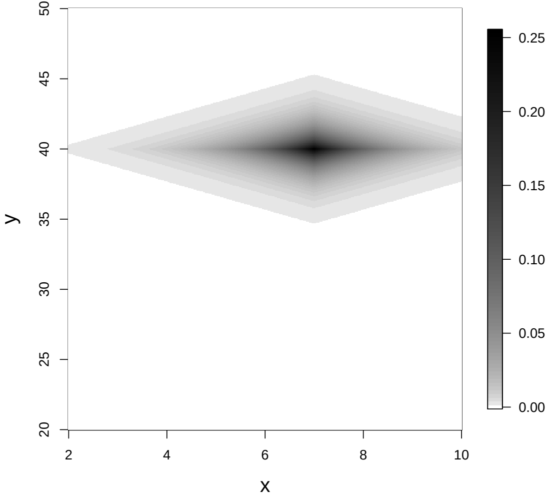
Continuous random variables

- Manipulating joint PDFs is similar to joint PMFs but sums are replaced by integrals
- The joint PDF is denoted $f(x, y)$
- Probabilities are computed as volume under the PDF:

$$\text{Prob}[(X, Y) \in A] = \int_A f(x, y)dxdy$$

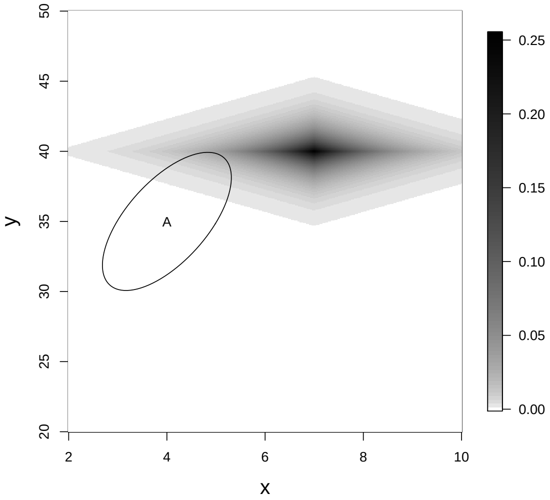
where $A \subset \mathcal{R}^2$

Joint PDF



The shade of a pixel is the joint PDF $f(x, y)$

Joint PDF



$P[(X, Y) \in A]$ is the integral under the surface over A

Continuous random variables

- Example: X=birthweight, Y=gestational age
- Domain: $X \in (2, 10)$ lbs and $Y \in (20, 50)$ weeks
- PDF: $f(x, y) = 0.26 \exp(-|x - 7| - |y - 40|)$
- Find: $\text{Prob}(X > 7, Y > 40)$

$$\text{Prob}(X > 7, Y > 40) =$$
$$\int_7^{10} \int_{40}^{50} 0.26 \exp(-|x - 7| - |y - 40|) dydx =$$
$$\cdots =$$
$$0.25$$

- See "BW Prob" in the online derivations

Continuous random variables

- The **Marginal PDF** of X is

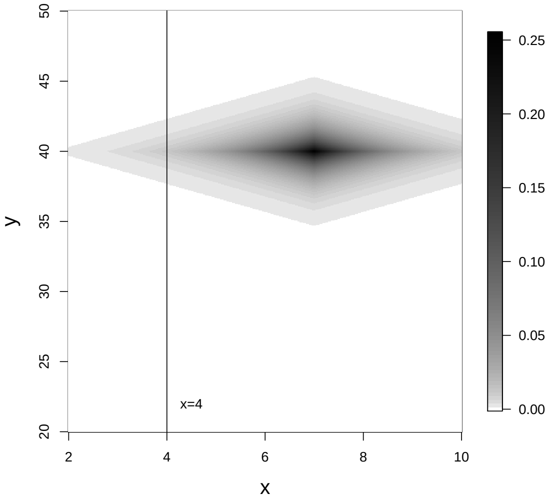
$$f_X(x) = \int f(x,y)dy$$

- f_X is the univariate PDF for X as if we never considered Y
- Find: $f_X(x)$ for the birthweight example

$$f_X(s) =$$
$$\int_{20}^{50} 0.26 \exp(-|x-7| - |y-40|)dydx =$$
$$\cdots =$$
$$0.52 \exp(-|x-7|)$$

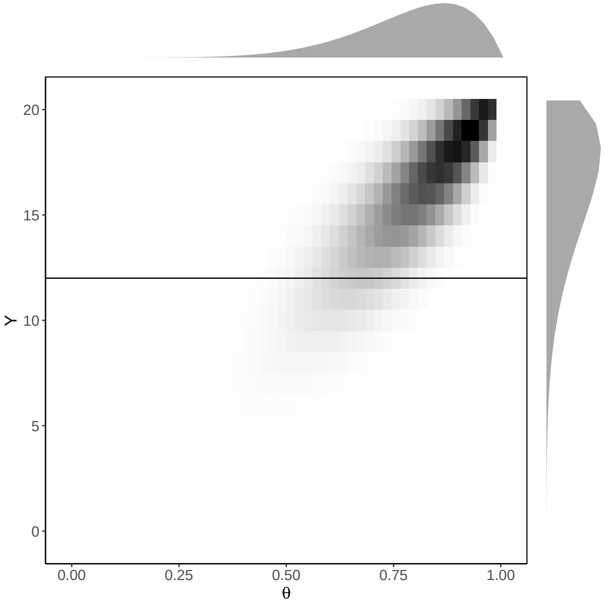
- See “BW marginal” in the online derivations

Marginal PDF



$f_X(4)$ is the integrated joint density along the line $x = 4$

Joint and marginal distributions



Continuous random variables

- The **Conditional PDF** of Y given X is

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

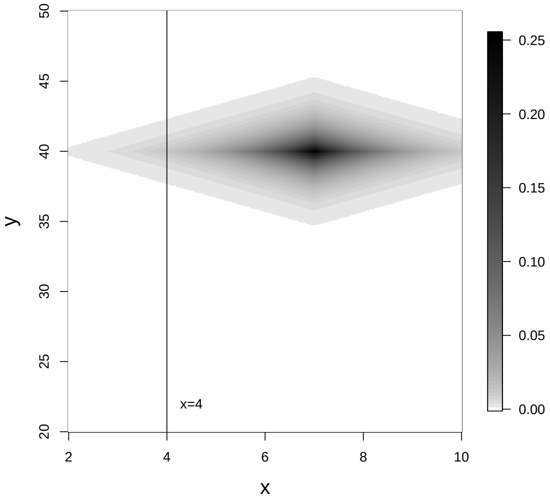
- Proper: $\int f(y|x)dy = \int \frac{f(x,y)}{f_X(x)}dy = \frac{\int f(x,y)dy}{f_X(x)} = 1$

- Find: $f(y|x)$ for the birthweight example

$$f(y|x) = \frac{0.26\exp(-|x-7| - |y-40|)}{0.52\exp(-|x-7|)} = 0.5\exp(-|y-40|)$$

- See "BW conditional" in the online derivations

Conditional PDF



$f(y|4)$ is the rescaled density along the line $x = 4$

Bivariate normal distribution

- The **bivariate normal distribution** is the most common multivariate family
- There are 5 parameters:
  - The marginal means of X and Y are μ_X and μ_Y
  - The marginal variances of X and Y are $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$
  - The correlation between X and Y is $\rho \in (-1, 1)$
- The joint PDF is $f(x, y) =$

$$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)}{2(1-\rho^2)}\right\}$$

Bivariate normal distribution

- The marginal distributions are $X \sim \text{Normal}(\mu_X, \sigma_X^2)$ and $Y \sim \text{Normal}(\mu_Y, \sigma_Y^2)$
- The conditional distribution is

$$Y|X \sim \text{Normal}\left\{\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X), \sigma_Y^2(1-\rho^2)\right\}$$

- The standard BVN distribution has $\mu_X = \mu_Y = 0$ and $\sigma_X = \sigma_Y = 1$ and thus

$$Y|X \sim \text{Normal}\left(\rho X, 1-\rho^2\right)$$

- Explain the role of ρ
- See "MVN marginal" and "MVN conditional" in the online derivations

Defining joint distributions conditionally

- Specifying joint distributions is hard
- Every joint distribution can be written

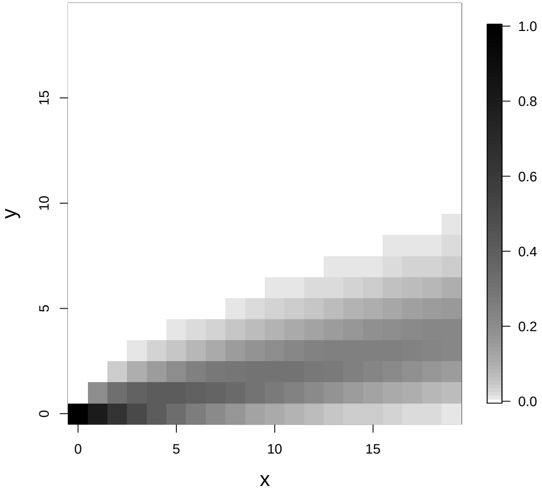
$$f(x, y) = f(y|x)f(x)$$

- Therefore, any joint distribution can be defined by
  1. X's marginal distribution
  2. The conditional distribution of $Y|X$
- The joint problem reduces to two univariate problems
- This idea forms the basis of hierarchical modeling

Defining joint distributions conditionally

- Let X be the number of robins in the forest
- Let Y be the number of robins we observe
- Model $\text{Prob}(X = x) = 1/20$ for $x \in \{0, ..., 19\}$ and $Y|X \sim \text{Binomial}(X, 0.2)$
- What is the support of (X, Y)?
- What is $\text{Prob}(X = 10, Y = 1)$?
- What is $\text{Prob}(Y = 0)$?

Joint PMF $f(x, y)$



$$f(x, y) = f(y|x)f_X(x) = \text{dbinom}(y, x, 0.2)(1/20)$$

Defining joint distributions conditionally

- What is the support of (X, Y)?

  The set of (x, y) so that $x, y \in \{0, 1, ..., 19\}$ and $y \leq x$ (the non-white pixels on the previous slide)

- What is Prob$(X = 10, Y = 1)$?

$$
\begin{aligned}
f(Y = 1|X = 10)f_X(10) &= \left[\binom{x}{y}0.2^y(1-0.2)^{x-y}\right][1/20] \\
&= 10 * 0.2^1 * 0.8^9 * 0.05 = 0.013
\end{aligned}
$$

Defining joint distributions conditionally

- What is Prob($Y = 0$)?

  The binomial($x, 0.2$) PMF at 0 is

$$f(Y = 0|x) = \binom{x}{0} 0.2^x 0.8^{x-0} = 0.8^x$$

  Therefore, the marginal probability Prob($Y = 0$) is

$$\begin{aligned} \sum_{x=0}^{19} f(x, 0) &= \sum_{x=0}^{19} f(0|x)(0.05) \\ &= \sum_{x=0}^{19} 0.8^x * 0.05 = 0.25 \end{aligned}$$