Marginal Distributions of Discrete Random Variables

Probability and Statistics for Data Science

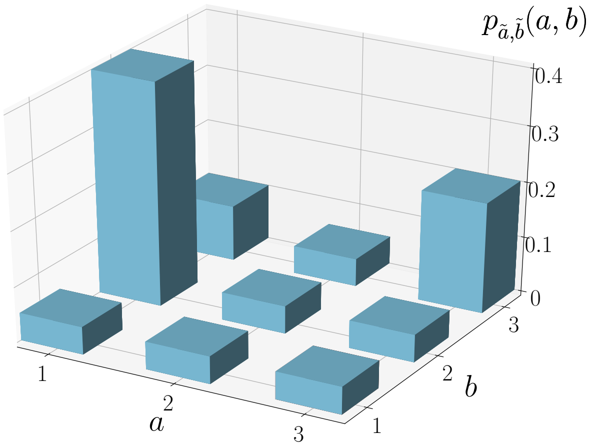
Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at `https://www.ps4ds.net`

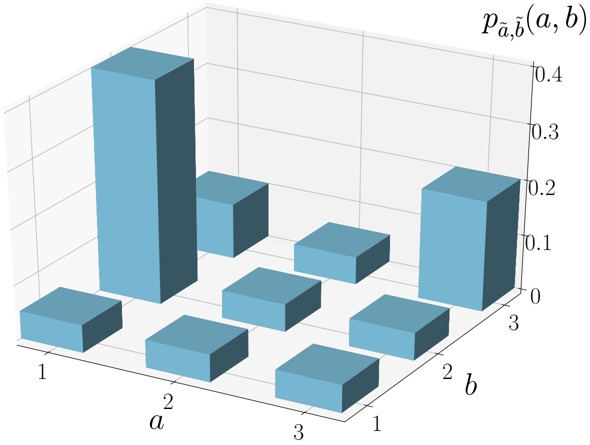
Motivation

In a model with many variables, how do we characterize behavior of individual variables?

Joint pmf



$p_{\tilde{a}}$?



$$p_{\tilde{a}}(1) = p_{\tilde{a},\tilde{b}}(1,1) + p_{\tilde{a},\tilde{b}}(1,2) + p_{\tilde{a},\tilde{b}}(1,3) = 0.55$$

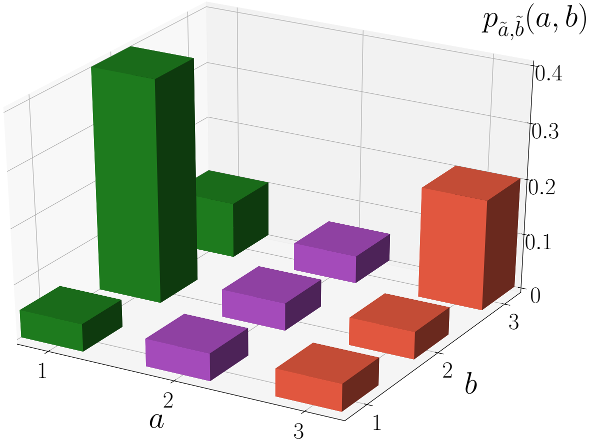
Marginal pmf

We have access to $p_{\tilde{a},\tilde{b}}$ but are only interested in $\tilde{a}$

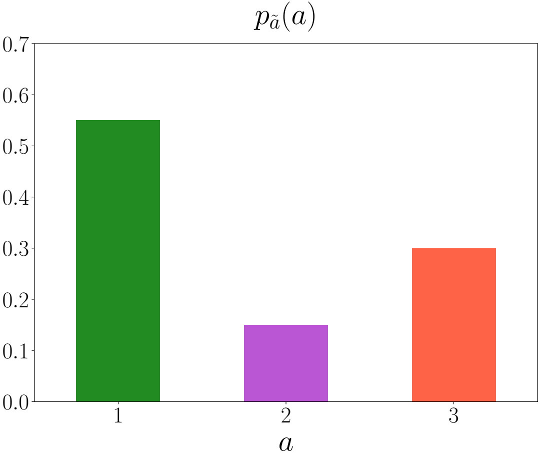
$$\begin{aligned}
p_{\tilde{a}}(a) &= \mathrm{P}(\tilde{a} = a) \\
&= \mathrm{P}\left(\cup_{b \in B} \left\{\tilde{a} = a, \tilde{b} = b\right\}\right) \\
&= \sum_{b \in B} \mathrm{P}\left(\tilde{a} = a, \tilde{b} = b\right) \\
&= \sum_{b \in B} p_{\tilde{a},\tilde{b}}(a, b)
\end{aligned}$$

The marginal pmf of $\tilde{a}$ is obtained by summing the joint pmf over all possible values of $\tilde{b}$

Marginal pmf



Marginal pmf



Movie ratings

Movielens dataset

Users give 1-5 ratings to movies

Goal: Model Independence Day and Mission Impossible

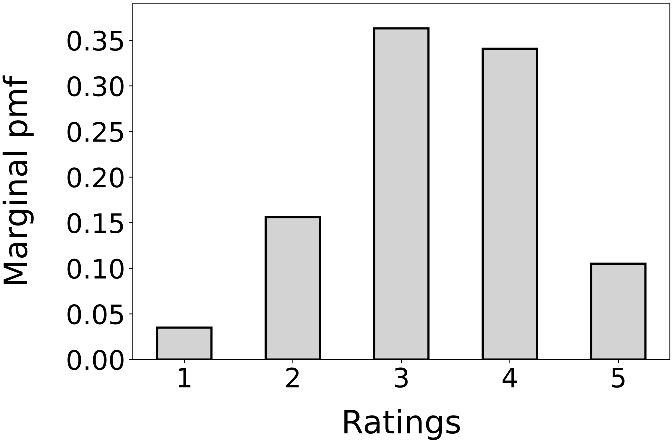
Empirical joint pmf (%)

		Independence Day				
		1	**2**	**3**	**4**	**5**
Mission Impossible	**1**	0.6	1	1.6	0.3	0
	2	1	3.8	5.7	3.5	1.6
	3	1.6	4.5	11.8	13.1	5.4
	4	1.9	4.8	6.4	15	6.1
	5	0	0	1.3	3.8	5.4

Mission Impossible?

		Independence Day				
		1	**2**	**3**	**4**	**5**
Mission Impossible	**1**	0.6	1	1.6	0.3	0
	2	1	3.8	5.7	3.5	1.6
	3	1.6	4.5	11.8	13.1	5.4
	4	1.9	4.8	6.4	15	6.1
	5	0	0	1.3	3.8	5.4

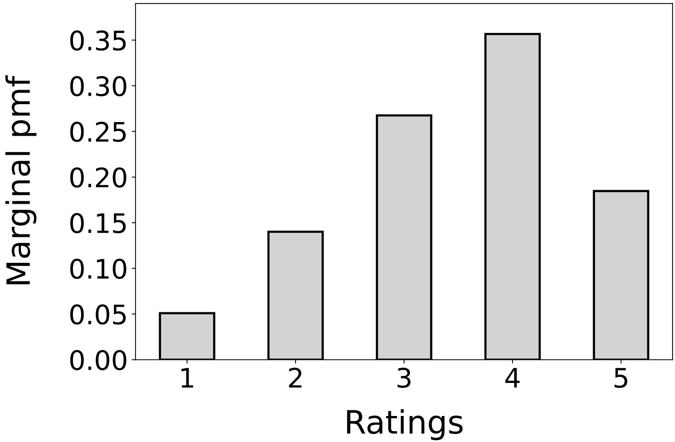
Marginal pmf



Independence Day?

		Independence Day				
		1	**2**	**3**	**4**	**5**
Mission Impossible	**1**	0.6	1	1.6	0.3	0
	2	1	3.8	5.7	3.5	1.6
	3	1.6	4.5	11.8	13.1	5.4
	4	1.9	4.8	6.4	15	6.1
	5	0	0	1.3	3.8	5.4

Marginal pmf

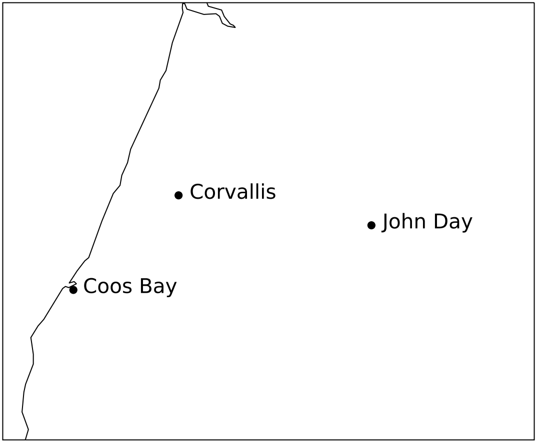


Marginal joint pmf

4-dimensional random vector $\tilde{x}$

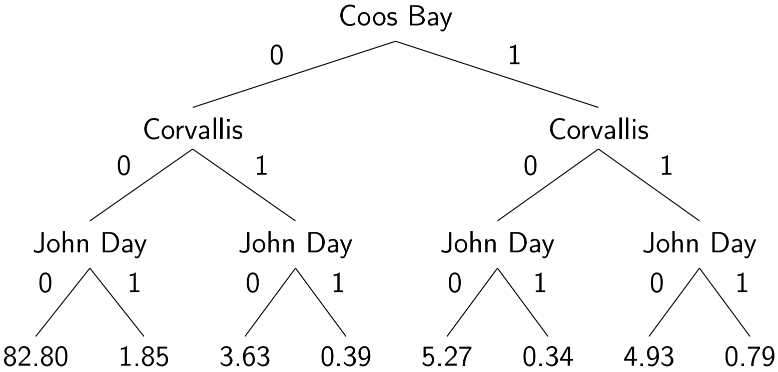
$$
\begin{aligned}
& p_{\tilde{x}[1],\tilde{x}[4]}(a,d) \\
& = \mathrm{P}\left(\cup_{b \in R_2, c \in R_3} \{\tilde{x}[1] = a, \tilde{x}[2] = b, \tilde{x}[3] = c, \tilde{x}[4] = d\}\right) \\
& = \sum_{b \in R_2} \sum_{c \in R_3} p_{\tilde{x}}(a,b,c,d)
\end{aligned}
$$

Precipitation in Oregon

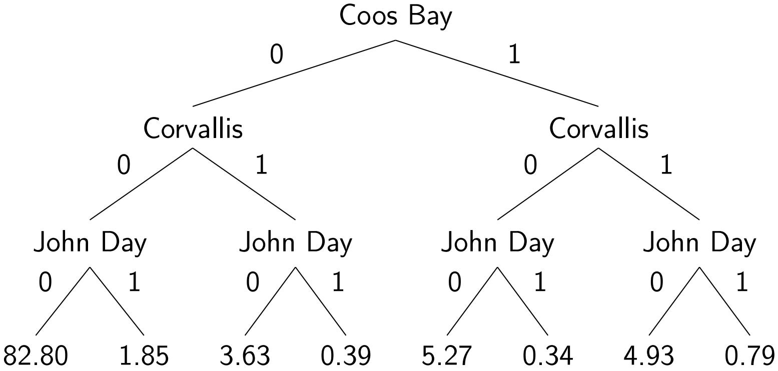


Goal: Model precipitation in Coos Bay, Corvallis, John Day

Precipitation in Oregon (%)



Coos Bay and Corvallis?



Coos Bay

0 1

Corvallis Corvallis

0 1 0 1

John Day John Day John Day John Day

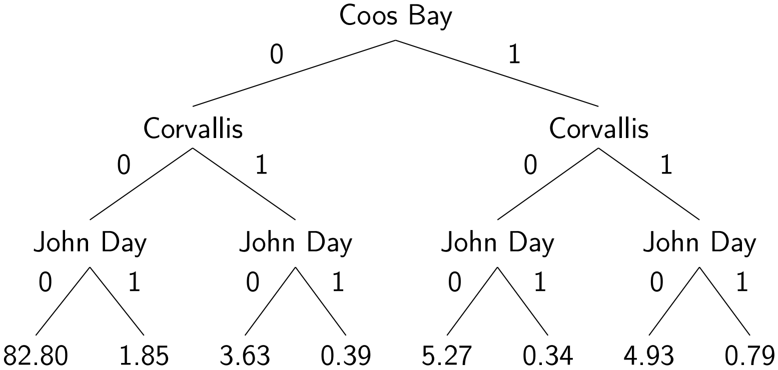
0 1 0 1 0 1 0 1

82.80 1.85 3.63 0.39 5.27 0.34 4.93 0.79

Marginal joint pmf

		Corvallis	
		0	**1**
Coos Bay	**0**	84.70	4.02
	1	5.62	5.72

Coos Bay and John Day?



Coos Bay

0 1

Corvallis Corvallis

0 1 0 1

John Day John Day John Day John Day

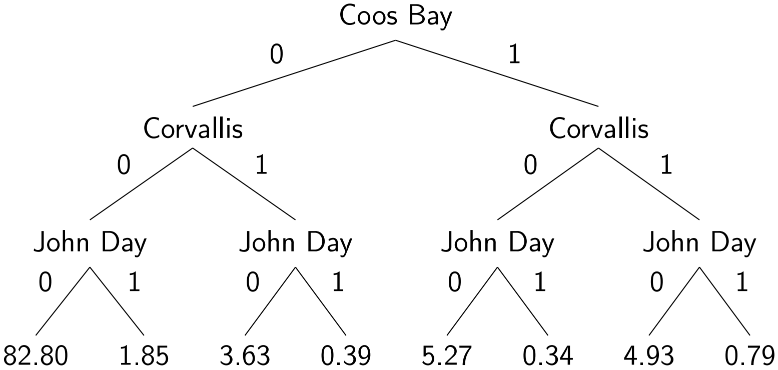
0 1 0 1 0 1 0 1

82.80 1.85 3.63 0.39 5.27 0.34 4.93 0.79

Marginal joint pmf

		John Day	
		0	**1**
Coos Bay	**0**	86.43	2.24
	1	10.21	1.13

Corvallis and John Day?



Coos Bay

0 1

Corvallis Corvallis

0 1 0 1

John Day John Day John Day John Day

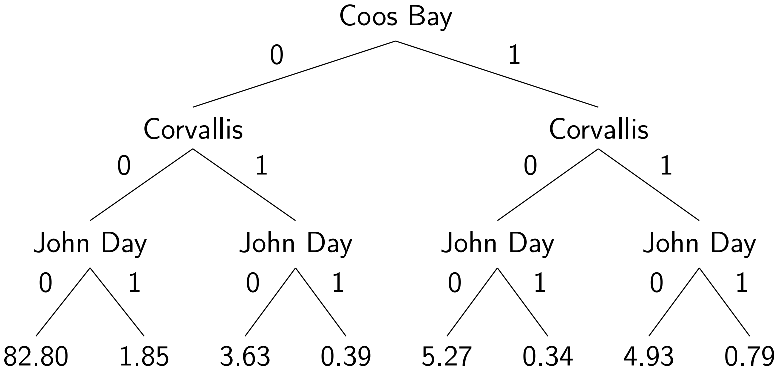
0 1 0 1 0 1 0 1

82.80 1.85 3.63 0.39 5.27 0.34 4.93 0.79

Marginal joint pmf

		John Day	
		0	**1**
Corvallis	**0**	88.07	2.19
	1	8.56	1.18

Coos Bay?



Coos Bay

0 1

Corvallis Corvallis

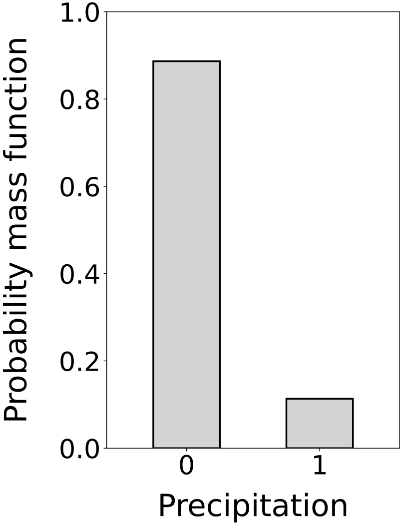
0 1 0 1

John Day John Day John Day John Day

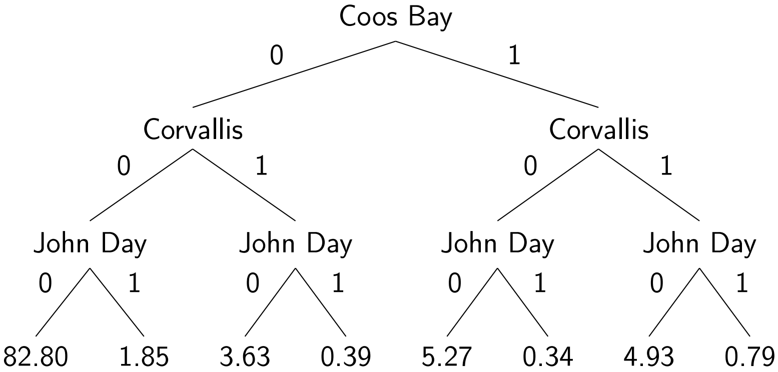
0 1 0 1 0 1 0 1

82.80 1.85 3.63 0.39 5.27 0.34 4.93 0.79

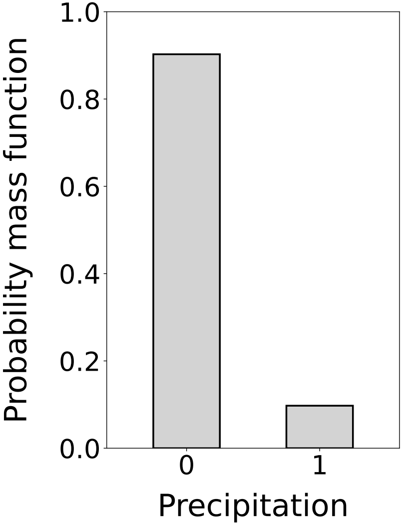
Marginal pmf



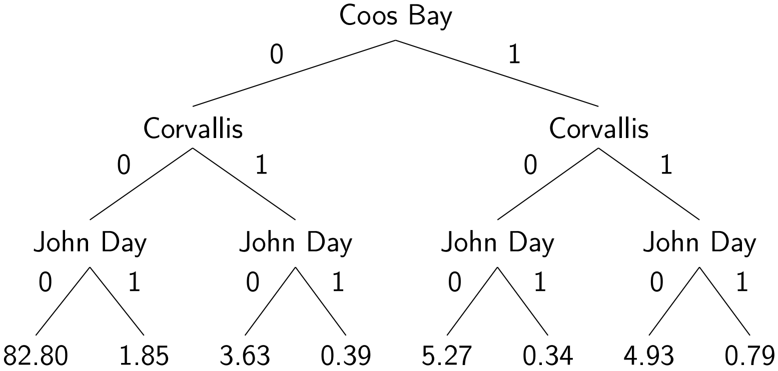
Corvallis?


Coos Bay
0
1
Corvallis
Corvallis
0
1
0
1
John Day
John Day
John Day
John Day
0
1
0
1
0
1
0
1
82.80
1.85
3.63
0.39
5.27
0.34
4.93
0.79

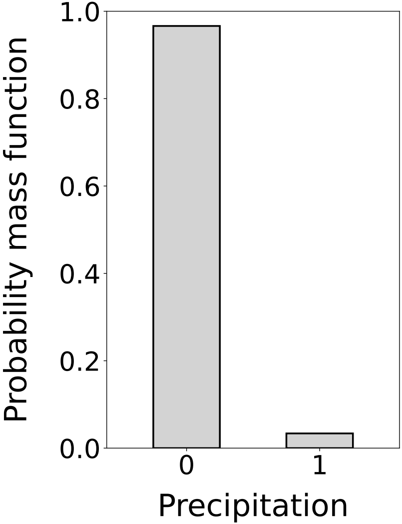
Marginal pmf



John Day?


Coos Bay
0
1
Corvallis
0
1
Corvallis
0
1
John Day
0
1
John Day
0
1
John Day
0
1
John Day
0
1
82.80
1.85
3.63
0.39
5.27
0.34
4.93
0.79

Marginal pmf



What have we learned?

How to compute marginal pmfs