Joint Distributions

We are often interested in the relationship between random variables
What is the relationship between winter snow depth and vegetation cover?
What is the relationship between race and mortgage lending?

Spatial Statistics:

What is the relationship between variables at different locations.
What is the relationship between my house price and my neighbor's?

The Joint CDF

Joint CDF of X and Y:
$F_{XY}(x, y) = P(X \leq x, Y \leq y)$

The Joint CDF

Joint CDF of X and Y:
$F_{XY}(x, y) = P(X \leq x, Y \leq y)$
Example: Raw Data

		T Training	No training	Total
A	Accident	30	70	100
	No Accident	570	330	900
	Total	600	400	1000

The Joint CDF

Joint CDF of X and Y:

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

Example: Recoding

		T		
		0	1	Total
A	0	30	70	100
	1	570	330	900
	Total	600	400	1000

The Joint CDF

Joint CDF of X and Y:
$F_{XY}(x,y) = P(X \leq x, Y \leq y)$
Example: Joint pmf $f(x,y) = P(X = x, Y = y)$

	T	
	0	1
A 0	.03	.07
1	.57	.33

The Joint CDF

Joint CDF of X and Y:
$F_{XY}(x,y) = P(X \leq x, Y \leq y)$
Example: Joint pmf $f(x,y) = P(X = x, Y = y)$

		T	
		0	1
A	0	.03	.07
	1	.57	.33

Example: Joint cdf

		T	
		0	1
A	0	.03	.1
	1	.6	1.00

Marginal pmfs

$f_X(x) = \sum_y f_{XY}(x,y) \left(= \int_y f_{XY}(x,y)dy \text{ for continuous RVs}\right)$

The marginal "collapses" the joint along one dimension

Example: joint pmf

		T	
		0	1
A	0	.03	.07
	1	.57	.33

Marginal pmfs

$f_X(x) = \sum_y f_{XY}(x,y) \left(= \int_y f_{XY}(x,y)dy \text{ for continuous RVs}\right)$

The marginal “collapses” the joint along one dimension

Example: Marginal pmf of A

		T		
		0	1	$f_A(a)$
A	0	.03	.07	.10
	1	.57	.33	.90

Marginal pmfs

$f_X(x) = \sum_y f_{XY}(x, y) \left(= \int_y f_{XY}(x, y)dy \text{ for continuous RVs}\right)$

The marginal "collapses" the joint along one dimension

Example: Marginal pmf of T

		T		
		0	1	
A	0	.03	.07	
	1	.57	.33	
	$f_T(t)$	.60	.40	

Conditional PMF

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

The conditional "slices" through a joint along one dimension

joint pmf

		T	
		0	1
A	0	.03	.07
	1	.57	.33

Conditional (on T) pmfs

		T	
		0	1
A	0	.05	.175
	1	.95	.825
		$f_{A\|T}(a\|T=0)$	$f_{A\|T}(a\|T=1)$

Conditional PMF

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

The conditional “slices” through a joint along one dimension

joint pmf

		T	
		0	1
A	0	.03	.07
	1	.57	.33

Conditional (on A) pmfs

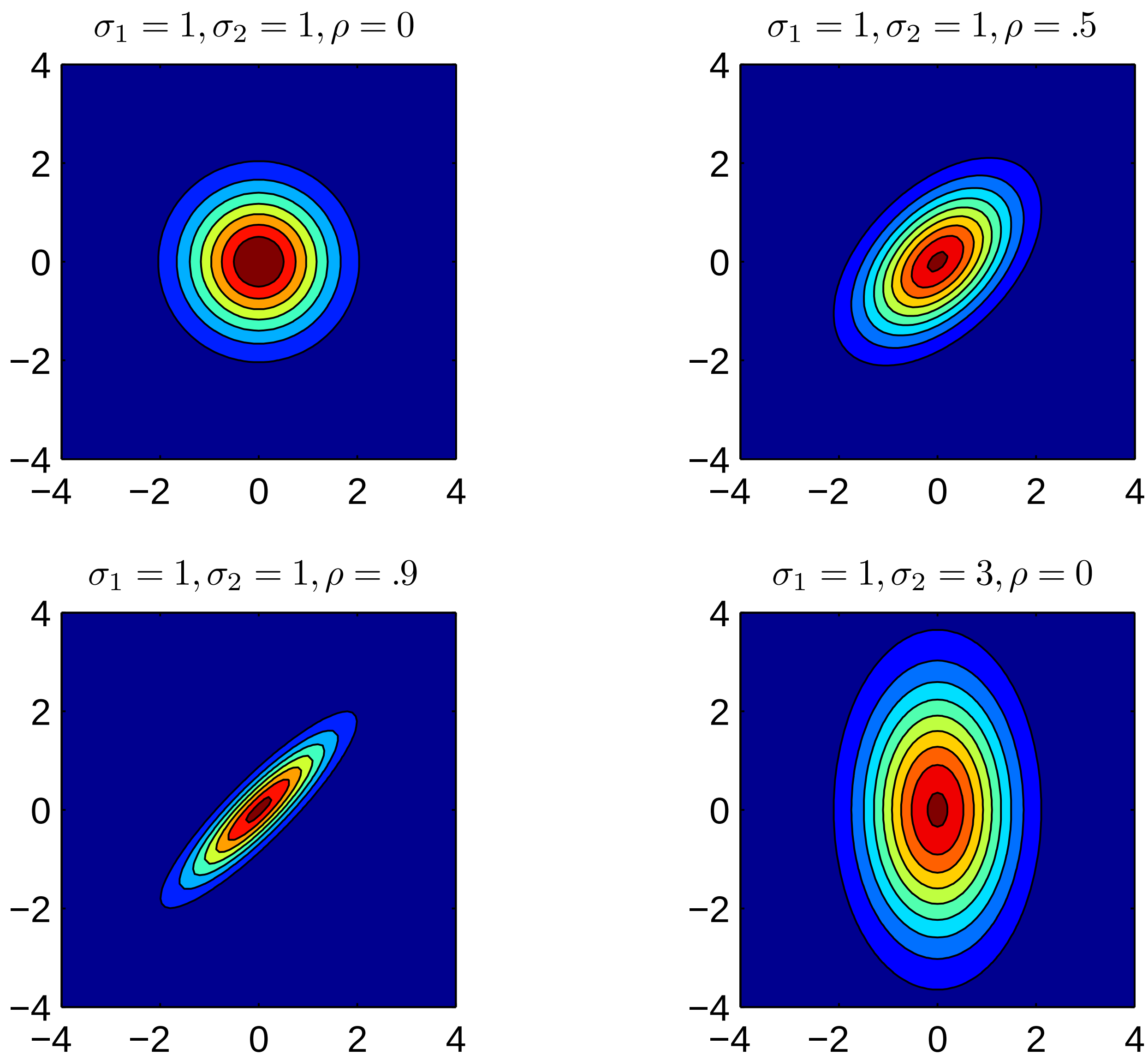
		T		
		0	1	
A	0	.30	.70	$f_{T\|A}(t\|A=0)$
	1	.633	.367	$f_{T\|A}(t\|A=1)$

Bivariate Normal Distribution

The bivareiate normal distribution is a bell curve in 2-D
Characterized by 5 numbers:

1. μ_X: Mean of X
2. μ_Y: Mean of Y
3. σ_X: Std. Deviation (scale) of X
4. σ_y: Std. Deviation (scale) of Y
5. ρ: Correlation of X and Y

Bivariate Normal Examples



Marginal and Conditional BVN

If $X, Y \sim BVN(\mu_x, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$

$$f_X(x) = N(\mu_x, \sigma_X^2)$$

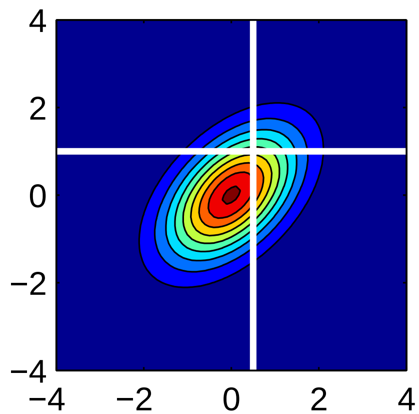
$$f_Y(y) = N(\mu_Y, \sigma_Y^2)$$

$$f_{X|Y}(X|Y = y) = N(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_y), \sigma_Y^2(1 - \rho^2))$$

Conditional Example

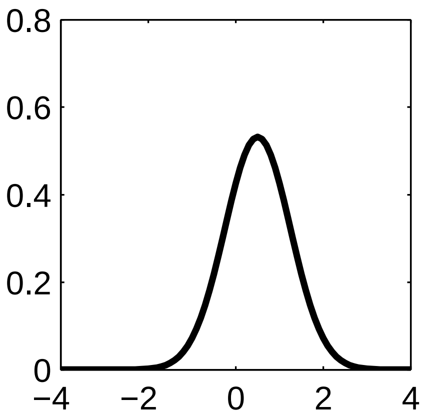
$\mu_x = \mu_y = 0 \; \sigma_x = \sigma_y = 1 \; \rho = .5$

Condition on $Y = 1$



$$\mu_{X|Y} = \mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

$$\mu_{X|Y} = 0 + .5 \frac{1}{1}(1 - 0) = .5$$



$$\sigma^2_{X|Y} = \sigma^2(1 - \rho^2)$$

$$\sigma^2_{X|Y} = 1(1 - .5^2) = .75$$