STSCI 4780/5780
Multivariate relationships, 2: The bivariate normal distribution and joint/conditional/marginal relationships

Tom Loredo, CCAPS & SDS, Cornell University

Recap

Building models for multivariate data via joint/conditional/marginal distribution relationships:

- Types of studies of multivariate data:
  - Correlation/dependence: Study joint, $p(x, y)$
  - Regression: Study conditional, $p(y|x)$
- BT as posterior = (joint for everything)/(marginal for knowns)
- Directed acyclic graphs (DAGs)
- Conditional independence
- Example: binomial prediction: $n_2 \perp\!\!\!\perp n_1|\alpha$

Agenda

1. **Bivariate normal distribution & regression**

2. **Joint from conditionals**

Agenda

❶ **Bivariate normal distribution & regression**

❷ **Joint from conditionals**

Regression perspective on the bivariate normal distribution

Outline of BVN development:

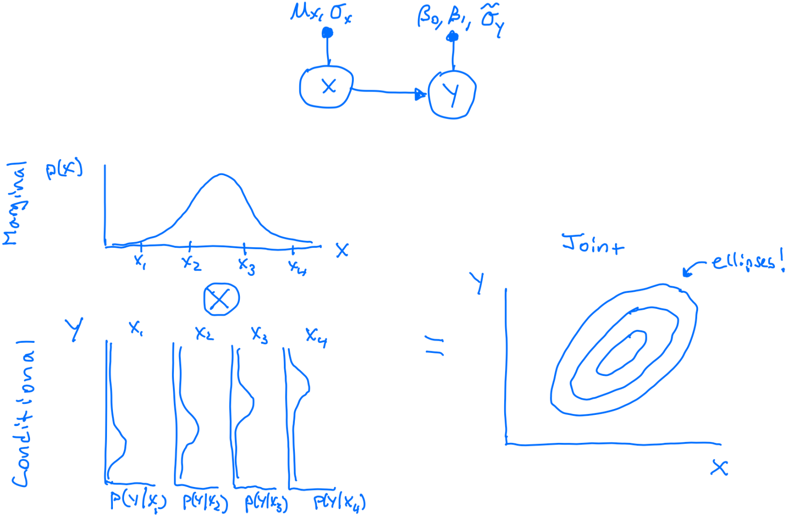
- Write a joint dist'n for two variables (x, y) factored as [marginal for x] $\times$ [cond'l for y given x]:
  - Normal marginal for x; mean μ_x, variance σ_x^2
  - Normal cond'l for y; conditional mean $\tilde{\mu}_y(x)$, conditional variance $\tilde{\sigma}_y^2$ (constant wrt x — homoskedastic)
  - Simplest nontrivial regression function:

$$\tilde{\mu}_y(x) \equiv \mathbb{E}(y|x) = \beta_0 + \beta_1 x$$

$$\Rightarrow \; p(x, y) = \mathrm{Norm}(x|\mu_x, \sigma_x) \times \mathrm{Norm}(y|\beta_0 + \beta_1 x, \tilde{\sigma}_y)$$

Five parameters: $\mu_x, \sigma_x; \; \beta_0, \beta_1, \tilde{\sigma}_y$

Visualizing $p(x,y) = \mathrm{Norm}(x|\mu_x, \sigma_x) \times \mathrm{Norm}(y|\beta_0 + \beta_1 x, \tilde{\sigma}_y)$:



$$
\begin{aligned}
p(x, y) &= p(x)\, p(y|x) \\
&\propto \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{[y-(\beta_0+\beta_1 x)]^2}{2\tilde{\sigma}_y^2}\right] \\
&= \exp\left[-\frac{Q(x, y)}{2}\right]
\end{aligned}
$$

with

$$
Q(x, y) = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{[y-(\beta_0+\beta_1 x)]^2}{\tilde{\sigma}_y^2}
$$

- Rewrite it as [marginal for y] $\times$ [cond'l for x given y]:
  - Find the marginal for y by completing the square in $x \rightarrow$ normal integral (after a surprising amount of algebra!)
  - Find the cond'l for x via $p(x, y) = p(y)p(x|y)$

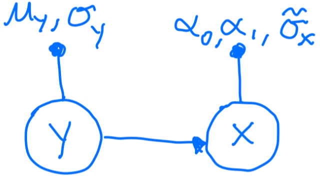
$$\rightarrow p(x|y) = \frac{p(x, y)}{p(y)}$$

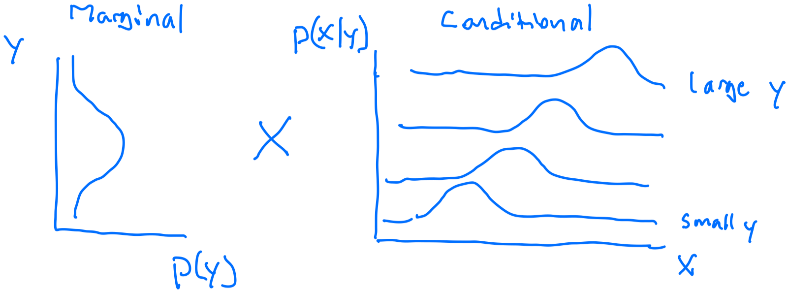
$$\Rightarrow \; p(x, y) = \mathrm{Norm}(y|\mu_y, \sigma_y) \times \mathrm{Norm}(x|\alpha_0 + \alpha_1 y, \tilde{\sigma}_x)$$

with

$$\mu_y = \beta_0 + \beta_1 \mu_x, \qquad \sigma_y^2 = \tilde{\sigma}_y^2 + \beta_1^2 \sigma_x^2;$$
$$\alpha_0 = \mu_x - \beta_1 \frac{\sigma_x^2}{\sigma_y^2} \mu_y, \qquad \alpha_1 = \beta_1 \frac{\sigma_x^2}{\sigma_y^2};$$
$$\tilde{\sigma}_x^2 = \sigma_x^2 \frac{\tilde{\sigma}_y^2}{\sigma_y^2}$$

Visualizing $p(x, y) = \mathrm{Norm}(y|\mu_y, \sigma_y) \times \mathrm{Norm}(x|\alpha_0 + \alpha_1 y, \tilde{\sigma}_y)$:





- Write the joint *symmetrically*, using the four marginal parameters $(\mu_x, \sigma_x, \mu_y, \sigma_y)$ and the *correlation coefficient*:

$$\rho = \beta_1 \frac{\sigma_x}{\sigma_y} = \alpha_1 \frac{\sigma_y}{\sigma_x}$$

$$\begin{aligned}
\Rightarrow \ p(x, y) &= C \exp\left[-\frac{Q(x, y)}{2}\right] \\
&= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{Q'(x, y)}{2(1-\rho^2)}\right]
\end{aligned}$$

$$Q'(x, y) = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho \, \frac{x-\mu_x}{\sigma_x} \, \frac{y-\mu_y}{\sigma_y}$$

with $|\rho| \leq 1$

- Regression functions in terms of joint parameters:

$$\begin{aligned}
\mathbb{E}(y|x) &= \mu_y + \beta_1(x - \mu_x) \\
&= \mu_y + \rho\sigma_y \left(\frac{x - \mu_x}{\sigma_x} \right) \\
\mathbb{E}(x|y) &= \mu_x + \alpha_1(y - \mu_y) \\
&= \mu_x + \rho\sigma_x \left(\frac{y - \mu_y}{\sigma_y} \right)
\end{aligned}$$

  The regression lines track the conditional mean, which is the (conditional) mode for normal distributions; thus they intersect contours of $p(x, y)$ where they are vertical (for $y|x$) or horizontal (for $x|y$)

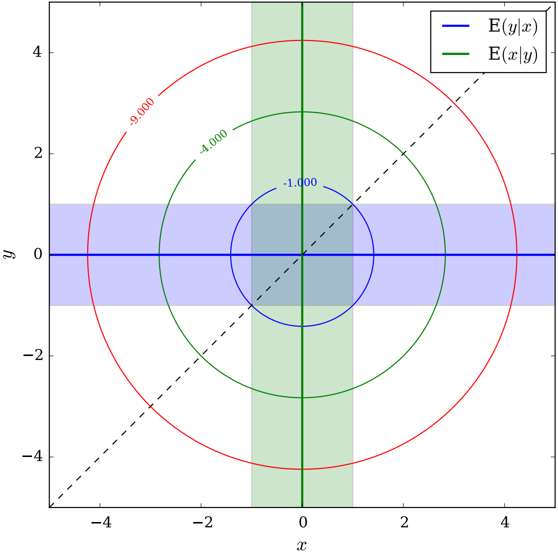
- Matrix forms (inverse covariance matrix vs. inverse correlation matrix):

$$
\begin{aligned}
Q'(x,y) &= \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\,\frac{x-\mu_x}{\sigma_x}\,\frac{y-\mu_y}{\sigma_y} \\
&= \begin{bmatrix}(x-\mu_x) & (y-\mu_y)\end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} \\
&= \begin{bmatrix}\frac{x-\mu_x}{\sigma_x} & \frac{y-\mu_y}{\sigma_y}\end{bmatrix} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} \frac{x-\mu_x}{\sigma_x} \\ \frac{y-\mu_y}{\sigma_y} \end{bmatrix}
\end{aligned}
$$

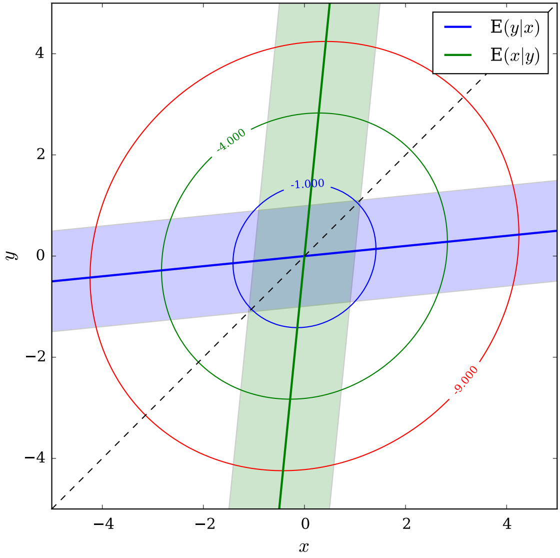
- Covariance and correlation matrices:

$$
\text{Covar} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}, \qquad \text{Corr} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}
$$

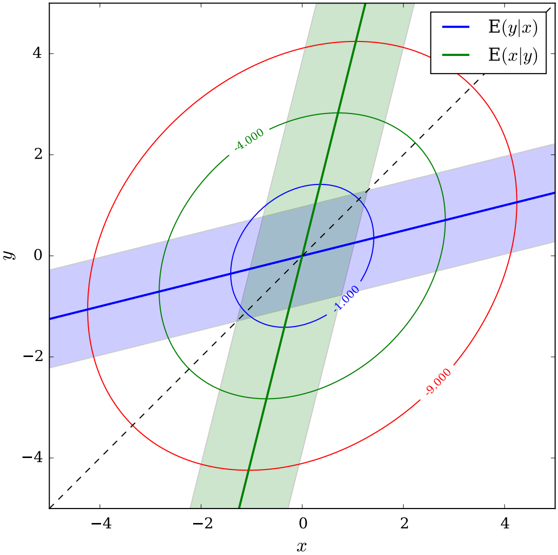
$\rho = 0$, $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$
Bands = regression curves $\pm\tilde{\sigma}_x, \tilde{\sigma}_y$
Contours = $p(x, y)$ level curves; labels show $\ln(p/p_{\mathrm{max}})$



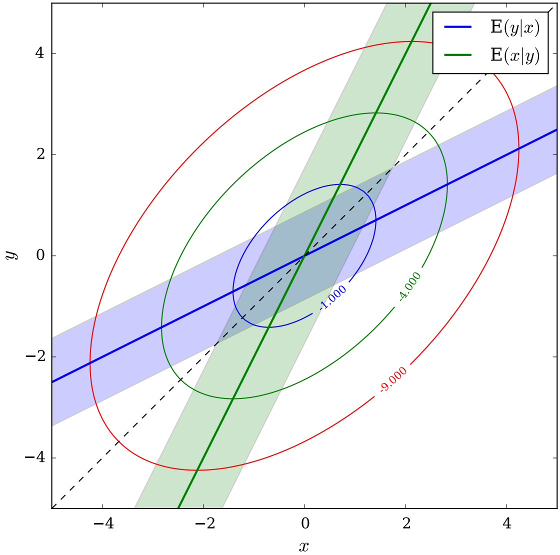
$\rho = 0.1$



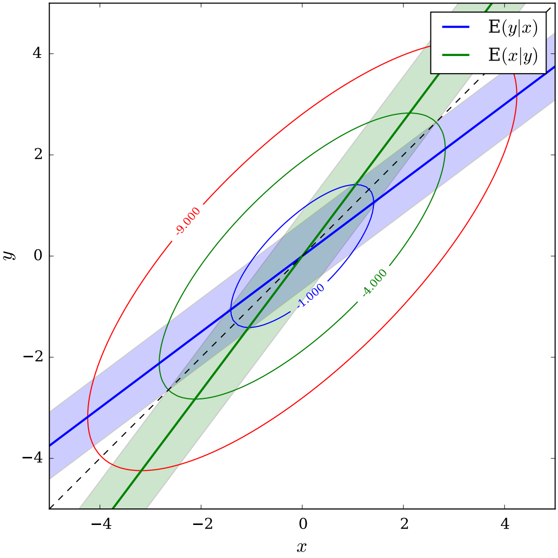
$\rho = 0.25$



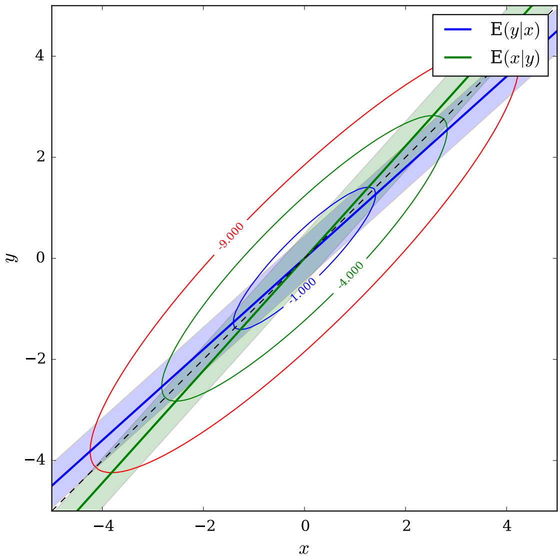
$\rho = 0.5$



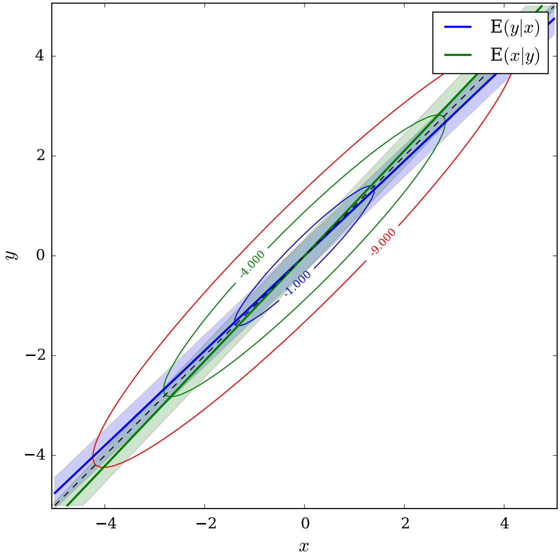
$\rho = 0.75$



$\rho = 0.9$



$\rho = 0.95$



Features to note

- The regression lines track the conditional mean, which is the (conditional) mode for normal distributions; thus they intersect contours where they are vertical (for $y|x$) or horizontal (for $x|y$)
- The regression lines for $y|x$ and $x|y$ are not the same; i.e., to find $\mathbb{E}(x|y)$ (a value of x as a function of y) we *do not* take the (blue) line:

  $$y(x) = \mathbb{E}(y|x) = \beta_0 + \beta_1 x$$

  and solve for $x(y) = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} y$
  In fact, the slopes of *both* regression lines are $\propto \beta_1$
- Neither regression line is the symmetry axis of the elliptical countours, except when $\rho = \pm 1$, when the ellipse collapses to a line
- The ellipse *always* has slope 1 (or -1), regardless of the value of ρ (or β_1); more generally its slope is σ_y / σ_x

Whence "regression"?

Recall the regression function for y given x

$$\mathbb{E}(y) \quad = \quad \mu_y + \rho\,\sigma_y\,\frac{x-\mu_x}{\sigma_x}$$

Compute *relative shifts from means*:

$$\rightarrow\ \frac{\mathbb{E}(y-\mu_y)}{\sigma_y} \quad = \quad \rho\,\frac{x-\mu_x}{\sigma_x}$$

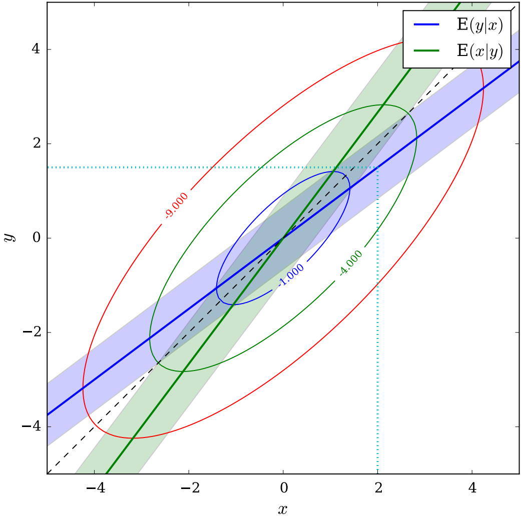
Since $|\rho| \leq 1$, so long as $\rho \neq 0$,

For a given (observed) x,
we expect y to deviate from its mean
by a smaller relative amount than x does from its mean

Galton, in a study of inheritance of stature, referred to this effect as "regression to mediocrity"

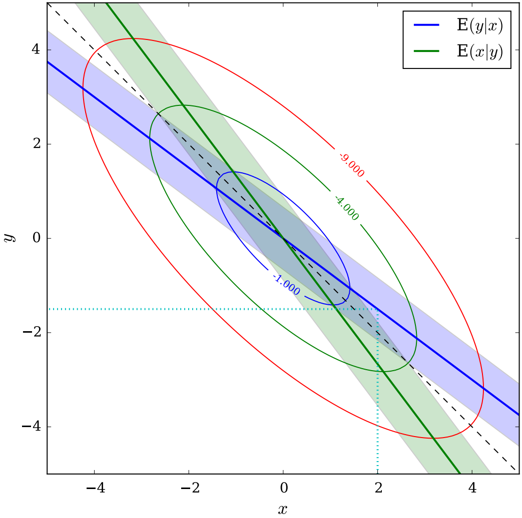
$\rho = 0.75,\ \mu_x = \mu_y = 0,\ \sigma_x = \sigma_y = 1$

$\mathbb{E}(y - \mu_y | x) < x - \mu_x$ when $x > \mu_x$ (else reverse)



Negative correlation: $\rho = -0.75$

$|\mathbb{E}(y - \mu_y|x)| < |x - \mu_x| \quad \text{when } x > \mu_x$



Agenda

1. Bivariate normal distribution & regression

2. Joint from conditionals

Joint distribution from conditionals?

The symmetric parameterization of the BVN has 5 parameters:

- Marginal means: μ_x, μ_y
- Marginal standard deviations: σ_x, σ_y
- Correlation coefficient: ρ

If we fix $(\mu_x, \sigma_x, \mu_y, \sigma_y)$ and vary ρ, we generate a family of distributions with *identical marginals but different joint distributions*

Specifying marginals does not *uniquely determine the joint*

Specifying one marginal and its associated conditional does give the joint:

$$\begin{aligned} p(x, y) &= p(x)\, p(y|x) \\ &= p(y)\, p(x|y) \end{aligned}$$

What about *specifying the two conditionals*?

Hammersley-Clifford theorem

We'll be evaluating joint, marginal, and conditional distributions for multiple choices of (x, y), so we introduce notation distinguishing the various functions (instead of using $p()$ for everything):

$$
\begin{aligned}
f(x, y) &\equiv p(x, y) \\
m_1(x) &\equiv p(x) = \int \mathrm{d}y\, p(x, y) \\
m_2(y) &\equiv p(y) = \int \mathrm{d}x\, p(x, y) \\
c_{12}(x; y) &\equiv p(x|y) \\
c_{21}(y; x) &\equiv p(y|x)
\end{aligned}
$$

From the product rule, for any choice of a, b,

$$\begin{aligned}
f(a,b) &= m_1(a)\, c_{21}(b;a) \\
\to m_1(a) &= \frac{f(a,b)}{c_{21}(b;a)}, \text{ for } \textit{any } b \\
\text{similarly } m_2(b) &= \frac{f(a,b)}{c_{12}(a;b)}, \text{ for } \textit{any } a
\end{aligned}$$

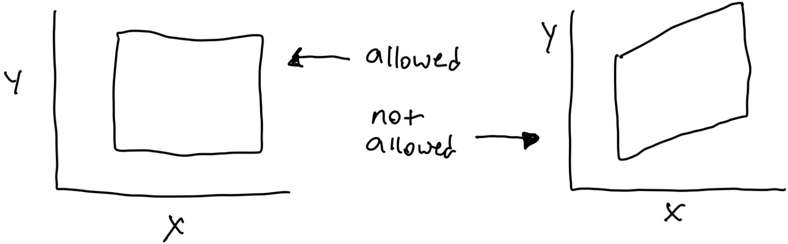
Now use the product rule for $p(x,y)$, replacing marginals:

$$\begin{aligned}
f(x,y) &= m_1(x)\, c_{21}(y;x) \\
&= \frac{f(x,b)}{c_{21}(b;x)}\, c_{21}(y;x), \text{ for any } b \\
&= \frac{m_2(b) c_{12}(x;b)}{c_{21}(b;x)}\, c_{21}(y;x) \\
&= f(a,b) \frac{c_{12}(x;b)}{c_{12}(a;b)}\, \frac{c_{21}(y;x)}{c_{21}(b;x)}
\end{aligned}$$

for any choice (a, b) (requires a *positivity condition*: support of joint = cartesian product of supports of marginals)

About the H-C positivity condition:

Box = support, where $p(x,y) > 0$



$$f(x,y) = f(a,b)\,\frac{c_{12}(x;b)}{c_{12}(a;b)}\,\frac{c_{21}(y;x)}{c_{21}(b;x)}$$

Here $f(a,b)$ is independent of (x,y), playing the role of a normalization constant for the remaining (x,y)-dependent factors

Knowing all the conditionals
uniquely determines the joint

Uses of this result (and its generalizations):

- Complex graphical models—Markov random fields
- Pseudo-likelihood methods
- *Gibbs sampling*: Using conditionals to build a MH proposal distribution