18.05 Lecture 19
March 28, 2005

Covariance and Correlation
Consider 2 random variables X, Y
$\sigma_x^2 = \mathrm{Var}(X), \sigma_y^2 = \mathrm{Var}(Y)$
Definition 1:
Covariance of X and Y is defined as:

$$\mathrm{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

Positive when both high or low in deviation.
Definition 2:
Correlation of X and Y is defined as:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

The scaling is thus removed from the covariance.

$\mathrm{Cov}(X, Y) = \mathbb{E}(XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y) =$
$= \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y - \mathbb{E}Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$$

Property 1:
If the variables are independent, $\mathrm{Cov}(X, Y) = 0$ (not correlated)
$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = \mathbb{E}X\mathbb{E}Y - \mathbb{E}X\mathbb{E}Y = 0$

Example: X takes values $\{-1, 0, 1\}$ with equal probabilities $\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$
$Y = X^2$
X and Y are dependent, but they are uncorrelated.
$\mathrm{Cov}(X, Y) = \mathbb{E}X^3 - \mathbb{E}X\mathbb{E}X^2$
but, $\mathbb{E}X = 0$, and $\mathbb{E}X^3 = \mathbb{E}X = 0$
Covariance is 0, but they are still dependent.
Also - Correlation is always between -1 and 1.

Cauchy-Schwartz Inequality:
$(\mathbb{E}XY)^2 \leq \mathbb{E}X^2\mathbb{E}Y^2$
Also known as the dot-product inequality:
$|(\vec{v}, \vec{u})| \leq \sqrt{|\vec{v}||\vec{u}|}$
To prove for expectations:

$$\phi(t) = \mathbb{E}(tX + Y)^2 = t^2\mathbb{E}X^2 + 2t\mathbb{E}XY + \mathbb{E}Y^2 \geq 0$$

Quadratic f(t), parabola always non-negative if no roots:
$D = (\mathbb{E}XY)^2 - \mathbb{E}X^2\mathbb{E}Y^2 \leq 0)$ (discriminant)
Equality is possible if $\phi(t) = 0$ for some point t.
$\phi(t) = \mathbb{E}(tX + Y)^2 = 0$, if tX + Y = 0, Y = -tX, linear dependence.
$(\mathrm{Cov}(X, Y))^2 = (\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y))^2 \leq \mathbb{E}(X - \mathbb{E}X)^2\mathbb{E}(Y - \mathbb{E}Y)^2 = \sigma_x^2\sigma_y^2$
$|\mathrm{Cov}(X, Y)| \leq \sigma_x\sigma_y$,

$$|\rho(X,Y)| = \frac{|\text{Cov}(X,Y)|}{\sigma_x \sigma_y} \leq 1$$

So, the correlation is between -1 and 1.

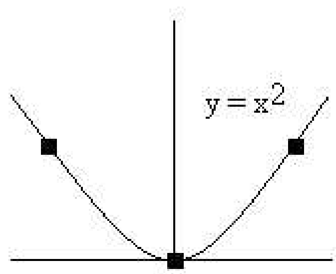
Property 2:

$$-1 \leq \rho(X,Y) \leq 1$$

When is the correlation equal to 1, -1?
$|\rho(X,Y)| = 1$ only when $Y - \mathbb{E}Y = c(X - \mathbb{E}X)$,
or $Y = aX + b$ for some constants a, b.
(Occurs when your data points are in a straight line.)
If $Y = aX + b$:

$$\rho(X,Y) = \frac{\mathbb{E}(aX^2 + bX) - \mathbb{E}X\mathbb{E}(aX+b)}{\sqrt{\text{Var}(X) \times a^2 \text{Var}(X)}} = \frac{a\text{Var}(X)}{|a|\text{Var}(X)} = \frac{a}{|a|} = sign(a)$$

If a is positive, then the correlation = 1, X and Y are completely positively correlated.
If a is negative, then correlation = -1, X and Y are completely negatively correlated.



Looking at the distribution of points on $Y = X^2$, there is NO linear dependence, correlation = 0.
However, if $Y = X^2 + cX$, then there is some linear dependence introduced in the skewed graph.

Property 3:

$$\text{Var}(X+Y) = \mathbb{E}(X + Y - \mathbb{E}X - \mathbb{E}Y)^2 = \mathbb{E}((X - \mathbb{E}X) + (Y - \mathbb{E}Y))^2 =$$

$$\mathbb{E}(X - \mathbb{E}X)^2 - 2\mathbb{E}(X - \mathbb{E}X)(\mathbb{E}(Y - \mathbb{E}Y) + \mathbb{E}(Y - \mathbb{E}Y)^2 = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y)$$

Conditional Expectation:
(X, Y) - random pair.
What is the average value of Y given that you know X?
f(x, y) - joint p.d.f. or p.f. then $f(y|x)$ - conditional p.d.f. or p.f.
Conditional expectation:

$$\mathbb{E}(Y|X=x) = \int yf(y|x)dy \text{ or } \sum yf(y|x)$$

$\mathbb{E}(Y|X) = h(X) = \int yf(y|X)dy$ - function of X, still a random variable.

Property 4:

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}Y$$

Proof:
$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(h(X)) = \int f(x)f(x)dx =$
$= \int(\int yf(y|x)dy)f(x)dx = \int\int yf(y|x)f(x)dydx = \int\int yf(x,y)dydx =$
$= \int y(\int f(x,y)dx)dy = \int yf(y)dy = \mathbb{E}Y$

Property 5:

$$\mathbb{E}(a(X)Y|X) = a(X)\mathbb{E}(Y|X)$$

See text for proof.

Summary of Common Distributions:

1) Bernoulli Distribution: $B(p), p \in [0,1]$ - parameter
Possible values of the random variable: $\mathcal{X} = \{0,1\}; f(x) = p^x(1-p)^{1-x}$
$\mathbb{P}(1) = p, \mathbb{P}(0) = 1-p$
$\mathbb{E}(X) = p, \text{Var}(X) = p(1-p)$

2) Binomial Distribution: $B(n,p)$, n repetitions of Bernoulli
$\mathcal{X} - \{0,1,...,n\}; f(x) = \binom{n}{x}p^x(1-p)^{1-x}$
$\mathbb{E}(X) = np, \text{Var}(X) = np(1-p)$

3) Exponential Distribution: $E(\alpha)$, parameter $\alpha > 0$
$\mathcal{X} = [0,\infty)$, p.d.f. $f(x) = \{\alpha e^{-\alpha x}, x \geq 0; 0, \text{ otherwise }\}$

$$\mathbb{E}X = \frac{1}{\alpha}, \mathbb{E}X^k = \frac{k!}{\alpha^k}$$

$$\text{Var}(X) = \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}$$

** End of Lecture 19