Lecture 5

Last time:
Characterizing groups of random variables
Names for groups of random variables

$$S = \sum_{i=1}^{n} X_i$$

$$\overline{S^2} = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{X_i X_j}$$

Characterize by pairs to compute

$$E[XY] = \overline{XY} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} xy f_{x,y}(x,y)dy$$

which we define as the *correlation*.

Often we do not know the complete distribution, but only simple statistics.

The most common of the moments of higher ordered distribution functions is the *covariance*,

$$\begin{aligned}\mu_{xy} &= E\left[(X-\bar{X})(Y-\bar{Y})\right] = \overline{(X-\bar{X})(Y-\bar{Y})} \\ &= \overline{XY} - \overline{X\bar{Y}} - \overline{\bar{X}Y} + \overline{\bar{X}\bar{Y}} \\ &= \overline{XY} - \bar{X}\bar{Y} - \bar{X}\bar{Y} + \bar{X}\bar{Y} \\ &= \overline{XY} - \bar{X}\bar{Y} \\ &= (\text{correlation}) - (\text{product of means})\end{aligned}$$

Even more significant is the *normalized covariance*, or *correlation coefficient*:

$$\rho = \frac{\mu_{xy}}{\sqrt{\sigma_x^2\sigma_y^2}} = \frac{\mu_{xy}}{\sigma_x\sigma_y}, \quad -1 \le \rho \le 1$$

This correlation coefficient may be thought of as measuring the <u>degree of linear dependence</u> between the random variables: $\rho = 0$ if the two are independent and $\rho = \pm 1$ if one is a linear function of the other. First note $\rho = 0$ if X and Y are <u>independent</u>.

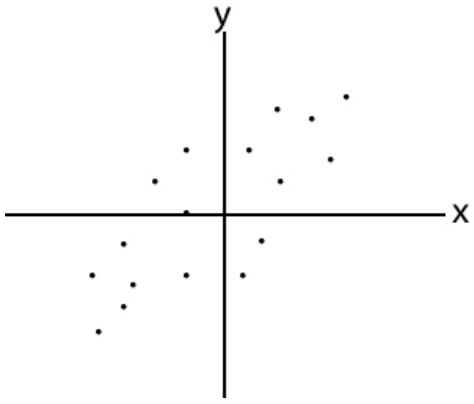
Calculate ρ_{xy} for $Y = a + bX$.

If <u>linearly related</u>:

$$\bar{Y} = a + b\bar{X}$$
$$\overline{XY} = a\bar{X} + b\overline{X^2}$$
$$\overline{Y^2} = a^2 + 2ab\bar{X} + b^2\overline{X^2}$$
$$\rho = \frac{a\bar{X} + b\overline{X^2} - \bar{X}(a + b\bar{X})}{\sqrt{\sigma_x^2(a^2 + 2ab\bar{X} + b^2\overline{X^2} - a^2 - 2ab\bar{X} - b^2\bar{X}^2)}}$$
$$= \frac{b\left(\overline{X^2} - \bar{X}^2\right)}{\sqrt{\sigma_x^2 b^2\left(\overline{X^2} - \bar{X}^2\right)}} = \frac{b\sigma_x^2}{\left|b\sigma_x^2\right|} = \pm 1 = \text{sgn}(b)$$

Degree of Linear Dependence

At every observation, or trial or the experiment, we observe a pair *x,y*. We ask: how well can we approximate *Y* as a linear function of *X*?



$$Y_{approx.} = a + bX$$

Choose *a* and *b* to minimize the mean squared error, $\overline{\varepsilon^2}$, in the approximation.

$$\varepsilon = Y_{approx.} - Y = a + bX - Y$$
$$\overline{\varepsilon^2} = \overline{a^2 + b^2X^2 + Y^2 + 2abX - 2bXY - 2aY}$$
$$= a^2 + b^2\overline{X^2} + \overline{Y^2} + 2ab\bar{X} - 2b\overline{XY} - 2a\bar{Y}$$

$$\frac{\partial \overline{\varepsilon^2}}{\partial a} = 2a + 2b\bar{X} - 2\bar{Y} = 0$$

$$\frac{\partial \overline{\varepsilon^2}}{\partial b} = 2b\overline{X^2} + 2a\bar{X} - 2\overline{XY} = 0$$

$$\frac{\partial \overline{\varepsilon^2}}{\partial b} - \bar{X}\frac{\partial \overline{\varepsilon^2}}{\partial a} = (2b\overline{X^2} - 2\overline{XY}) - (2b\bar{X}^2 - 2\bar{X}\bar{Y}) = 0$$

$$b = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - \bar{X}^2} = \frac{\mu_{xy}}{\sigma_x^2} = \rho\frac{\sigma_y}{\sigma_x}$$

$$a = \bar{Y} - \frac{\mu_{xy}}{\sigma_x^2}\bar{X} = \bar{Y} - \rho\frac{\sigma_y}{\sigma_x}\bar{X}$$

$$Y_{approx.} = \bar{Y} - \frac{\mu_{xy}}{\sigma_x^2}\bar{X} + \frac{\mu_{xy}}{\sigma_x^2}X$$

$$= \bar{Y} + \frac{\mu_{xy}}{\sigma_x^2}\left(X - \bar{X}\right)$$

$$= \bar{Y} + \rho\frac{\sigma_y}{\sigma_x}\left(X - \bar{X}\right)$$

$$\varepsilon = Y_{approx.} - Y = \bar{Y} + \rho\frac{\sigma_y}{\sigma_x}\left(X - \bar{X}\right) - Y$$

$$\bar{\varepsilon} = 0$$

$$\overline{\varepsilon^2} = a(a + 2b\bar{X} - 2\bar{Y}) + b^2\overline{X^2} + \overline{Y^2} - 2b\overline{XY}$$

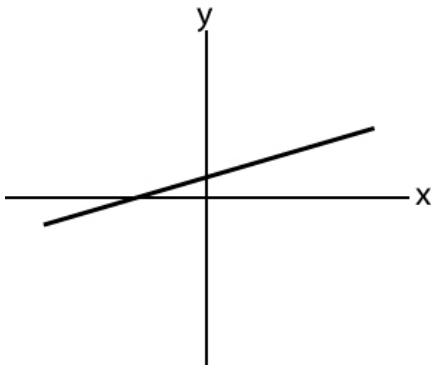
$$= (\bar{Y} - \frac{\mu_{xy}}{\sigma_x^2}\bar{X})(\bar{Y} - \frac{\mu_{xy}}{\sigma_x^2}\bar{X} + 2\frac{\mu_{xy}}{\sigma_x^2}\bar{X} - 2\bar{Y}) + \frac{\mu_{xy}}{\sigma_x^2}\overline{X^2} + \overline{Y^2} - 2\frac{\mu_{xy}}{\sigma_x^2}\overline{XY}$$

$$= -\bar{Y}^2 + \frac{\mu_{xy}}{\sigma_x^2}\bar{X}\bar{Y} + \frac{\mu_{xy}}{\sigma_x^2}\bar{X}\bar{Y} - \frac{\mu_{xy}}{\sigma_x^2}\bar{X}^2 + \frac{\mu_{xy}}{\sigma_x^2}\overline{X^2} + \overline{Y^2} - 2\frac{\mu_{xy}}{\sigma_x^2}\overline{XY}$$

$$= \sigma_y^2 - 2\frac{\mu_{xy}}{\sigma_x^2}\mu_{xy} + \frac{{\mu_{xy}}^2}{\sigma_x^4}\sigma_x^2$$

$$= \sigma_y^2 - \frac{{\mu_{xy}}^2}{\sigma_x^2} = \sigma_y^2\left(1 - \frac{{\mu_{xy}}^2}{\sigma_x^2\sigma_y^2}\right)$$

$$= \sigma_y^2(1 - \rho^2)$$



If X and Y were actually linearly related, the points would appear on one straight line, ρ would be ± 1, and the mean squared error in the approximation would be zero.

If X and Y were independent, the points would scatter all over the x,y plane, μ_{xy} would be zero, so $Y_{approx.} = \bar{Y}$, and $\overline{\varepsilon^2} = \sigma_y^2$.

Note that dependence other than linear is not necessarily measured by ρ.

> ***Example*: $Y = X^2$ and $\bar{X} = \overline{X^3} = 0$.**
> $\mu_{xy} = \overline{XY} = \bar{X}\bar{Y} = \overline{X^3} - \bar{X}\,\overline{X^2} = 0$
> $\rightarrow \rho = 0,$ but X, Y are dependent!

Also, high correlation does not imply cause and effect.

> ***Example: Dying in the hospital.***
> A survey reports that two events, "entering the hospital" and "dying within 1 week" have a high correlation. This relationship, however, is not causal. There exists a third, unreported event, "disease," which causes each of the other events.

Vector-Matrix Notation

Define the vectors $\underline{X}$ and $\underline{x}$ and the mean $E[\underline{X}]$.

$$E[\underline{X}] = \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n \underline{x} f_n(\underline{x})$$

$$= \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} dx_n \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} f_n(\underline{x})$$

$$= \underline{\bar{X}}$$

$$= \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{bmatrix}$$

Correlation Matrix

$$E\left[\underline{X}\underline{X}^T\right] = \overline{\underline{X}\underline{X}^T} = M$$

$$M_{ij} = \overline{X_i X_j}$$

So the correlation matrix arrays all the correlations among the X_i with the mean squared values on the diagonal.

Note that:

- The correlation matrix is symmetric.
- If X_i and X_j are independent, the correlation is the product of the means (from the product rule for the pdf).
  $\overline{X_i X_j} = \bar{X}_i \bar{X}_j$

Covariance Matrix

$$E\left[(\underline{X} - \underline{\bar{X}})(\underline{X} - \underline{\bar{X}})^T\right] = E\left[\underline{X}\underline{X}^T\right] - E[\underline{X}]\underline{\bar{X}}^T - \underline{\bar{X}}E\left[\underline{X}^T\right] + \underline{\bar{X}}\underline{\bar{X}}^T$$

$$= E\left[\underline{X}\underline{X}^T\right] - \underline{\bar{X}}\underline{\bar{X}}^T - \underline{\bar{X}}\underline{\bar{X}}^T + \underline{\bar{X}}\underline{\bar{X}}^T$$

$$= \overline{\underline{X}\underline{X}^T} - \underline{\bar{X}}\underline{\bar{X}}^T \equiv C$$

$$C_{ij} = \left[(\underline{X} - \underline{\bar{X}})(\underline{X} - \underline{\bar{X}})^T\right]_{ij}$$

$$= \overline{X_i X_j} - \bar{X}_i \bar{X}_j$$

This is by definition the covariance between X_i and X_j. So the covariance matrix arrays all the covariances among the X_i with the variances along the diagonal.

If X_i and X_j are independent, their covariance is zero.

If the covariance between two variables is zero, they are said to be uncorrelated. This does not, in general, imply that they are independent.

Conditional Distribution

Sometimes the notion of a conditional distribution is important. If we wish to confine our attention to the subject of cases in which an event E occurs, we would define the *conditional probability distribution function*:

$$F_E(x) = F(x \mid E) = P(X \le x \mid E)$$
$$= \frac{P(X \le x, E)}{P(E)}$$

and the *conditional probability density function*:

$$f_E(x) = f(x \mid E) = \frac{dF_E(x)}{dx}$$

Application of Bayes' Rule

Bayes' theorem can be written for the distribution of a random variable conditioned on the occurrence of an event E or the observation of a certain value of another random variable.

Conditioned on an event E:

Original form: $P(A_k \mid E) = \dfrac{P(A_k)P(E \mid A_k)}{\sum_i P(A_i)P(E \mid A_i)}$

Let the events $A_i \rightarrow$ the events $x < X \le x + dx$

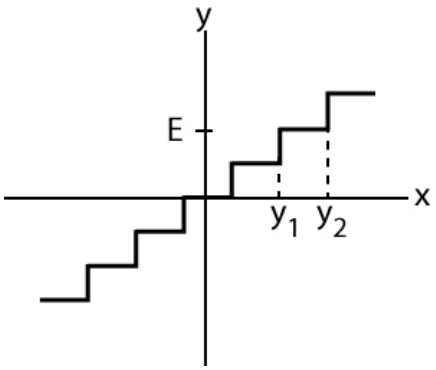
Note that for different values x as $dx \rightarrow 0$ these events are mutually exclusive and collectively exhaustive.

$$\lim_{dx \to 0} f(x \mid E)dx = \frac{f(x)dxP(E \mid x)}{\int_{-\infty}^{\infty} f(u)P(E \mid u)du}$$

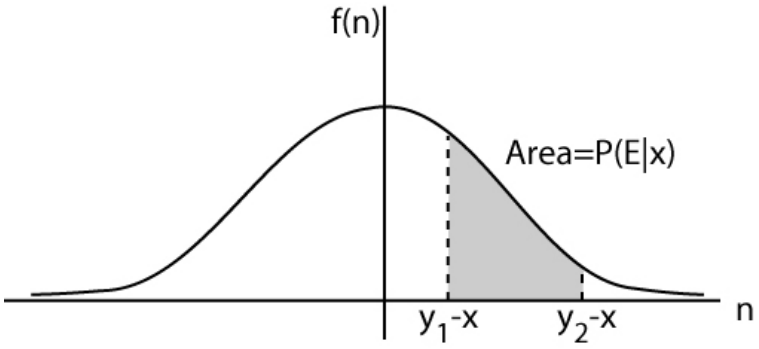
$$f(x \mid E) = \frac{f(x)P(E \mid x)}{\int_{-\infty}^{\infty} f(u)P(E \mid u)du}$$

Example: Measuring the position of a spacecraft far away

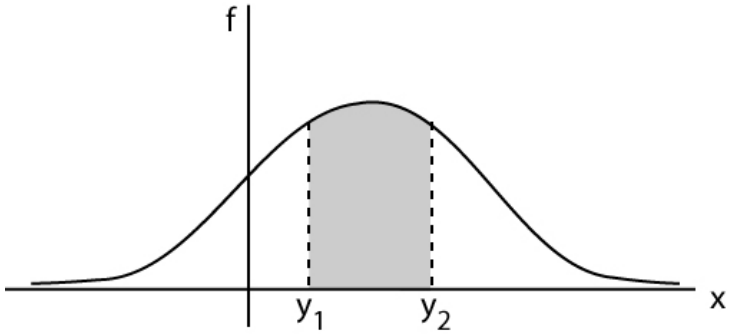
Measurement: $y = x + n$, where n is noise. E implies y lies in a quantized interval. It is simple to use this interval to estimate y, but not correct.



$$E \Rightarrow y_1 < y \le y_2$$
$$y_1 < x + n \le y_2$$
$$y_1 - x < n \le y_2 - x$$



We update the unconditional distribution for x to the conditional distribution for x. This is the right way.



Conditioned on a value of Y:
Let the event E be the event $y < Y \le y + dy$. E will be the value taken by y (like a measurement variable).

$$\lim_{dy \to 0} f(x \mid y) = \frac{f(x) f(y \mid x) dy}{\int_{-\infty}^{\infty} f(u) f(y \mid u) dy du}$$

Note: y is not the variable of integration.

$$f(x \mid y) = \frac{f(x) f(y \mid x)}{\int_{-\infty}^{\infty} f(u) f(y \mid u) du}$$

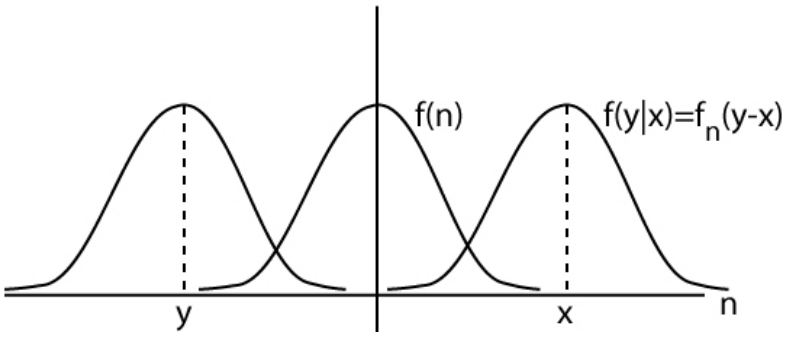
This provides a basis for the estimation of a parameter X based on prior information, $f(x)$, and measurements y which are related to X.

Example: $y = x + n \quad n = y - x$ ***parameterizing*** x

$f(y \mid x) = f_n(y - x)$

We assume n is independent of x.

$f(y \mid x)$ is the noise distribution shifted by x.



Conditional Expectation

The conditional expectation of a random variable or function of a random variable is the expectation calculated with the conditional distribution.

$$E[g(x) \mid A] = \int_{-\infty}^{\infty} g(x) f(x \mid A) dx$$

A useful relation involving conditional expectations is

$$E[g(x)] = \sum_i P(A_i) E[g(X) \mid A_i]$$

if the A_i are mutually exclusive and exhaustive. Or, if we let the events A_i be the events $y < Y \le y + dy$,

$$E[g(x)] = \int_{-\infty}^{\infty} E[g(x) \mid y] f(y) dy$$

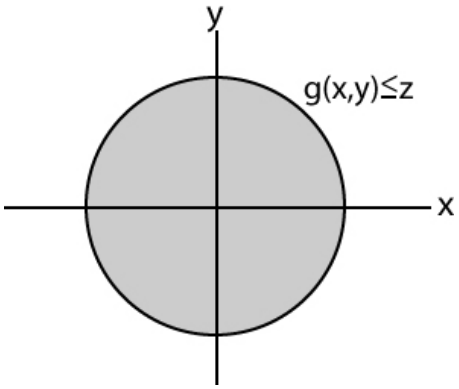
The conditional expectation of a random variable, $E[x \mid y]$, is often taken as the estimate of the random variable after incorporating the effect of a measured event or related random variable.

$f(x \mid E_1, ..., E_n)$: State of knowledge about *x* given observations $E_1, ..., E_n$

$f(x \mid y_1, ..., y_n)$: State of knowledge about *x* given measurements $y_1, ..., y_n$

Probability Distribution of Functions of Random Variables

The probability distribution and density for functions of random variables can be calculated from the distributions of the variables themselves.



In general, if $Z = g(X,Y)$ and we want $f_z(z)$, we can use

$$\begin{aligned} F_z(z) &= P(Z \le z) \\ &= P[g(X,Y) \le z] \\ &= \int dx \int dy f_{x,y}(x,y) \end{aligned}$$

to get the distribution function for all values of *z*.

$$f_z(z) = \frac{dF_z(z)}{dz}$$

In particular cases, there may be easier ways to do this, but this is the general procedure.