Lecture 3

Last time: Use of Bayes' rule to find the probability of each outcome in a set of

$$P(A_k \mid E_1, E_2, ...E_n)$$

In the special case when E's are conditionally independent (though they all depend on the alternative, A$_k$),

$$P(A_k \mid E_1, E_2, ...E_n) = \frac{P(A_k \mid E_1...E_{n-1})P(E_n \mid A)}{\sum_i P(A_i \mid E_1...E_{n-1})P(E_n \mid A_i)}.$$

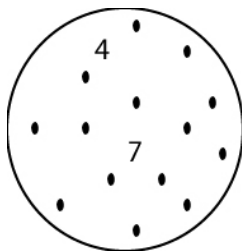This is easy to do and can be done recursively.

Take E_1: $$P(A_k \mid E_1) = \frac{P(A_k)P(E_1 \mid A_k)}{\sum_i P(A_i)P(E_1 \mid A_i)}$$

Take E_2: $$P(A_k \mid E_1 E_2) = \frac{P(A_k \mid E_1)P(E_2 \mid A_k)}{\sum_i P(A_i \mid E_1)P(E_2 \mid A_i)}$$
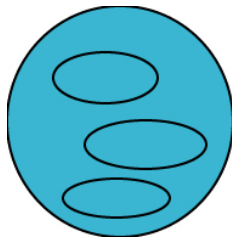
➜ <u>Recursive application</u> of Bayes' rule.

<u>Random Variables</u>

Start with a conceptual <u>random experiment</u>:

➜ Discrete-valued random variable

➜ Continuous-valued random variable

The *random variable* is <u>defined to be a function</u> of the outcomes of the random experiment. Example: **X** = the number of spots on the top face of a die.

How to <u>characterize</u> a random variable?

Probability distribution function (or cumulative probability function)

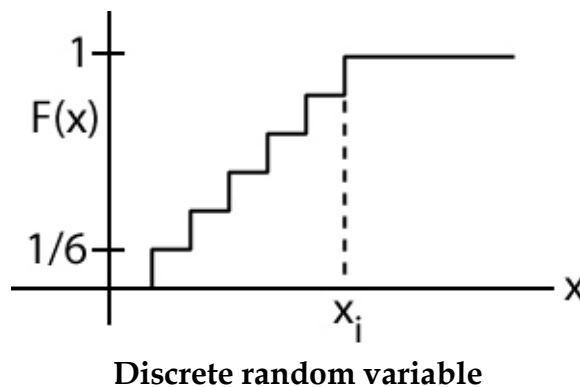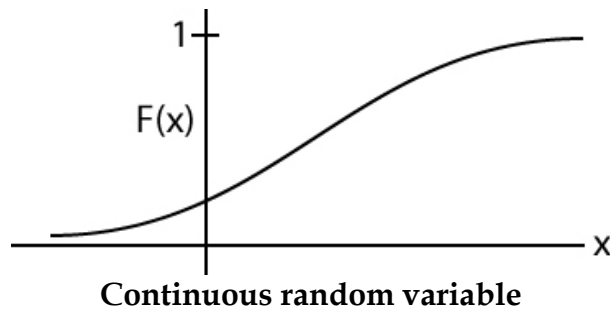$F(x) = P(X \leq x)$, where x is the *argument* and X is the *random variable name*.

Properties:
$F(-\infty) = 0$
$F(\infty) = 1$
$0 \leq F(x) \leq 1$
$F(b) \geq F(a)$, if b>a

Continuous random variable

Discrete random variable

An <u>alternate way</u> to characterize random variables is the *probability density function*:

$f(x) = \dfrac{dF(x)}{dx}$

$f(x) \geq 0, \forall x$

$f(x) = F(-\infty) + \displaystyle\int_{-\infty}^{x} f(u)du$

$\quad = \displaystyle\int_{-\infty}^{x} f(u)du$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$
$$= F(b) - F(a)$$
$$= \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$
$$= \int_{a}^{b} f(x)dx$$

$$F(\infty) = \int_{-\infty}^{\infty} f(u)du = 1$$

The density function is always a non-negative function, which integrates over the full range to 1.

Consider the probability that an observation lies in a narrow interval between x and $x+dx$.

$$\lim_{dx \to 0} P(x < X \leq x + dx) = \lim_{dx \to 0} \int_{x}^{x+dx} f(u)du$$
$$= \lim_{dx \to 0} f(x)dx$$

This can be used to measure $f(x)$.

We see clearly from this that the probability of a random variable having a continuous distribution function taking any prescribed value ($dx \to 0$) is zero.
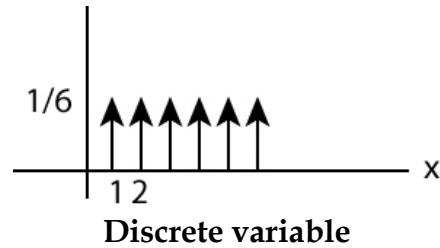
In the case of a continuous random variable, the probability that it takes any exact value is 0.
Also we shall use this relation in setting up problem solutions to indicate the probability of x taking a value in an infinitesimal interval near x.

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}$$

The appropriate range has to be chosen due to the fact that if you make intervals too small you'll have too few samples lying in each interval. The sampling error is too great if the number of samples you get in each interval is too small. If you make the interval too large then the approximation made in the limit is no longer applicable.
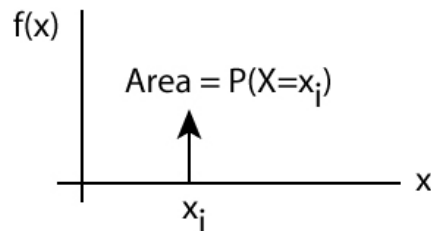
Discrete variable

Discrete variable

Using the Dirac delta function

$$\delta(x) = 0, x \neq 0$$

$$\int_{0^-}^{0^+} \delta(x)dx = 1$$

The function $f(x) = \delta(x - x_i)$

we have

$$f(x) = \sum_i P(X = x_i)\delta(x - x_i)$$

Expection

The <u>expectation</u> or <u>expected value</u> of a function of a random variable is defined as:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \equiv \text{mean value } \overline{X}$$

$$= \sum_i x_i P(X = x_i) \text{ for a discrete-valued random variable } X$$

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

$$= \sum_i g(x_i)P(X = x_i) \text{ for a discrete-valued random variable } X$$

Some common expectations which partially characterize a random variable- moments of the distribution:

1st: Mean or mean value

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \bar{X}, \text{ deterministic}$$

2nd: Mean squared value

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \overline{X^2}$$

These are the <u>first two moments</u> of the distribution. The complete infinite set of moments would completely characterize the random variable.

We also define <u>central moments</u> (moments about the mean). The first non-zero one is:
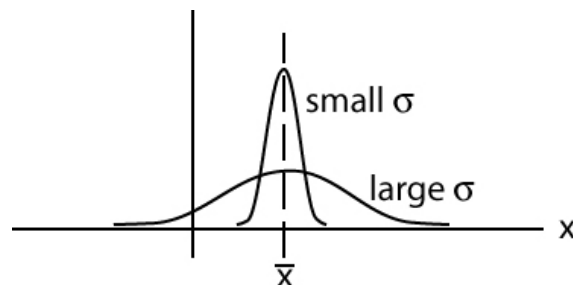
1st: Variance

$$E[(X - \bar{X})^2] = \int_{-\infty}^{\infty} (x - \bar{X})^2 f(x) dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - 2\bar{X} \int_{-\infty}^{\infty} x f(x) dx + \bar{X}^2 \int_{-\infty}^{\infty} f(x) dx$$

$$= \overline{X^2} - 2\bar{X}^2 + \bar{X}^2$$

$$= \overline{X^2} - \bar{X}^2 \equiv \sigma^2$$

Standard Deviation

$\sigma = \sqrt{\text{Variance}}$

$\bar{X}$ indicates the center of the distribution

σ indicates the width of the distribution

The variance of a random variable is the mean squared value of its deviation from its mean. In some sense then, the variance must measure the probability of stated deviations from the mean.

A very general quantitative statement of this is known as <u>Chebyshev's</u>
<u>Inequality</u>:

$$P(|X - \bar{X}| \geq t) \leq \frac{\sigma^2}{t^2}$$

This is a very general result in that it holds for any distribution of *X*.

Note that it gives you just an upper bound on the probability of exceeding a
given deviation from the mean. Being complete general, we should not expect
this bound to be a tight one quantitatively. Because of its generality, it is a
powerful analytical tool. For quantitative practical work it is preferable if
possible to estimate the distribution of *X* and use this to calculate the probability
indicated above.

Characteristic Function

1. Definition of Characteristic Function
2. Inverse relation
3. Characteristic function and inverse for discrete distributions
4. Prove $\phi_x(t) = \Pi \phi_{x_i}(t)$ where $X = X_1 + X_2 + ... + X_n$ (X_i independent)
5. MacLaurin series expansion of $\phi(t)$

1. Definition of Characteristic Function

$$\phi(t) = E(e^{jtx})$$

$$= \int_{-\infty}^{\infty} e^{jtx} f_x(x) dx$$

The characteristic function of a random variable *X* with density function $f_x(x)$ is
defined to be the expectation of e^{jtx}; the integral expression then follows from the
relation for the expectation of a function of a random variable.

The main purpose in defining the characteristic function is its utility in the
treatment of many problems.

2. Inverse relation

The expression for $\phi(t)$ is in the form of the inverse Fourier transform of the
density function fo X; thus by the theory of Fourier transforms the direct relation
must hold and the functions involved are uniquely related. Note that the
integral always converges.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jxt} \phi(t) dt$$

So *f(x)* and *ϕ(t)* imply the same information.

3. Characteristic function and inverse for discrete distributions

For discrete distributions, the density function is the sum of terms of the type

$$f_i(x) = p_i \delta(x - x_i)$$

For each such term, the characteristic function has an additive term of the form

$$\phi_i(t) = \int_{-\infty}^{\infty} e^{jtx} f_i(x) dx$$

$$= \int_{-\infty}^{\infty} e^{jtx} p_i \delta(x - x_i) dx$$

$$= p_i e^{jtx_i}$$

If the inverse relation is to make sense in this case we must have

$$f_i(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jxt} \phi_i(t) dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jxt} p_i e^{jtx_i} dt$$

$$= p_i \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jt(x-x_i)} dt = p_i \delta(x - x_i)$$

This requires

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jt(x-x_i)} dt = \delta(x - x_i),$$

which is true of any use of a delta function in transform operations. This is also true for the opposite transform pair, where the exponent of *e* is positive.