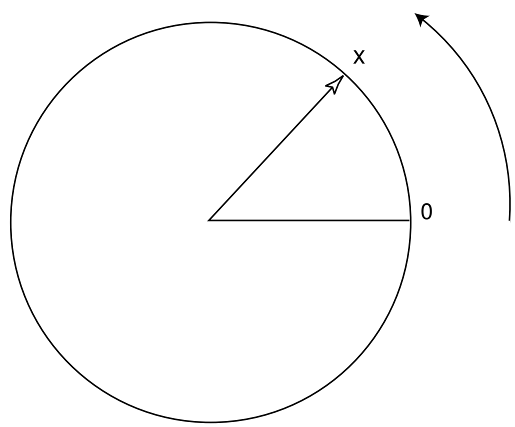
Central Limit Theorem

7/24/2006

Continuous Probability Densities

- Let us construct a spinner, which consists of a circle of unit circumference and a pointer.



- The experiment consists of spinning the pointer and recording the label of the point at the tip of the pointer.

- We let the random variable X denote the value of this outcome.
- The sample space is clearly the interval $[0, 1)$.
- It is necessary to assign the probability 0 to each outcome.
- The probability

$$P(0 \leq X \leq 1)$$

should be equal to 1.

- We would like the equation

$$P(c \leq X < d) = d - c$$

  to be true for every choice of c and d.

- If we let $E = [c, d]$, then we can write the above formula in the form

$$P(E) = \int_E f(x)\, dx \ ,$$

  where $f(x)$ is the constant function with value 1.

Density Functions of Continuous Random Variables

Let X be a continuous real-valued random variable. A *density function* for X is a real-valued function f which satisfies

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

for all $a,\ b \in \mathbf{R}$.

- It is *not* the case that all continuous real-valued random variables possess density functions.

- In terms of the density $f(x)$, if E is a subset of $\mathbb{R}$, then

$$P(X \in E) = \int_E f(x)\, dx \ .$$

Example

- In the spinner experiment, we choose for our set of outcomes the interval $0 \le x < 1$, and for our density function

$$f(x) = \begin{cases} 1, & \text{if } 0 \le x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

- If E is the event that the head of the spinner falls in the upper half of the circle, then $E = \{\, x : 0 \le x \le 1/2 \,\}$, and so

$$P(E) = \int_0^{1/2} 1 \, dx = \frac{1}{2} \; .$$

- More generally, if E is the event that the head falls in the interval $[c, d]$, then

$$P(E) = \int_c^d 1\,dx = d - c\ .$$

Example: Continuous Uniform Density

- The simplest density function corresponds to the random variable U whose value represents the outcome of the experiment consisting of choosing a real number at random from the interval $[a, b]$.

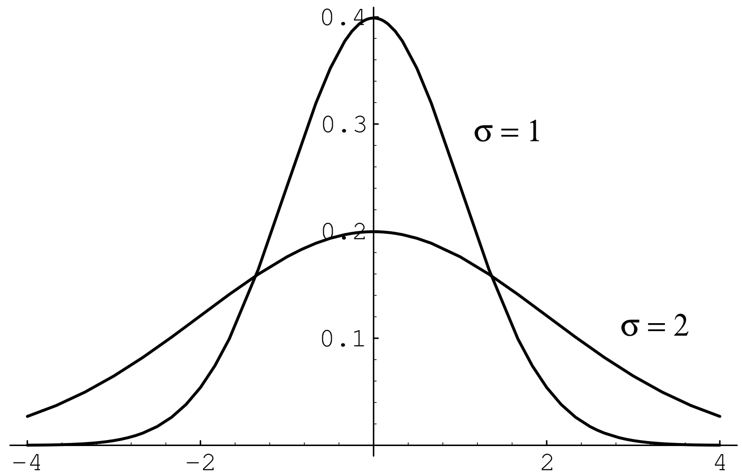
$$f(x) = \begin{cases} 1/(b-a), & \text{if } a \le x \le b \\ 0, & \text{otherwise.} \end{cases}$$

Normal Density

- The *normal density* function with parameters μ and σ is defined as follows:
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \ .$$

- The parameter μ represents the "center" of the density.

- The parameter σ is a measure of the "spread" of the density, and thus it is assumed to be positive.


0.4
0.3
0.2
0.1
σ = 1
σ = 2
-4
-2
2
4

Central Limit Theorem for Bernoulli Trials

- We deal only with the case that $\mu = 0$ and $\sigma = 1$.

- We will call this particular normal density function the *standard* normal density, and we will denote it by $\phi(x)$:

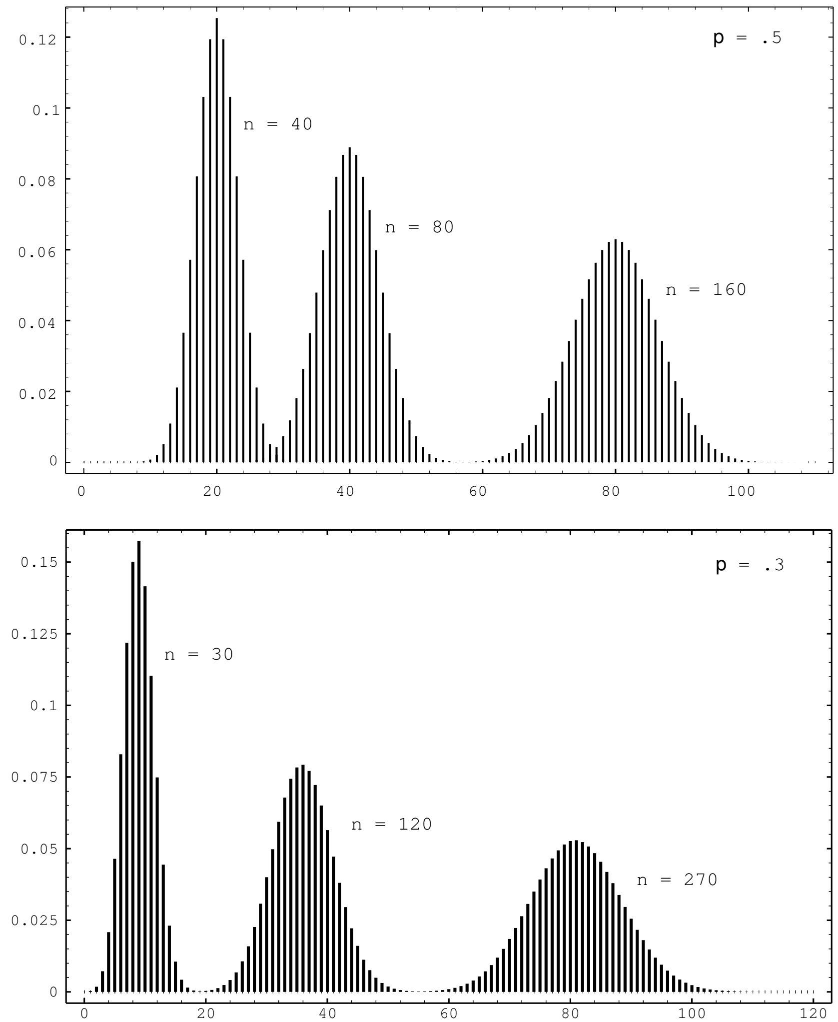
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ .$$

- Consider a Bernoulli trials process with probability p for success on each trial.

- Let $X_i = 1$ or 0 according as the ith outcome is a success or failure, and let $S_n = X_1 + X_2 + \cdots + X_n$.

- Then S_n is the number of successes in n trials.

- We know that S_n has as its distribution the binomial probabilities $b(n, p, j)$.


p = .5
n = 40
n = 80
n = 160
p = .3
n = 30
n = 120
n = 270

Standardized Sums

- We can prevent the drifting of these spike graphs by subtracting the expected number of successes np from S_n.
- We obtain the new random variable $S_n - np$.
- Now the maximum values of the distributions will always be near 0.
- To prevent the spreading of these spike graphs, we can normalize $S_n - np$ to have variance 1 by dividing by its standard deviation $\sqrt{npq}$

Definition

The *standardized sum* of S_n is given by

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \ .$$

S_n^* always has expected value 0 and variance 1.

- We plot a spike graph with the spikes placed at the possible values of S_n^*: x_0, x_1, $\ldots$, x_n, where

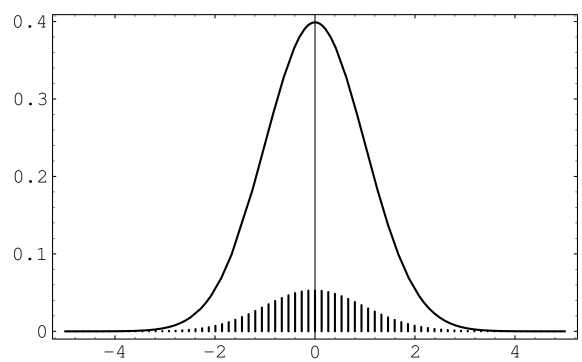
$$x_j = \frac{j - np}{\sqrt{npq}} \ .$$

- We make the height of the spike at x_j equal to the distribution value $b(n, p, j)$.

- We plot a spike graph with the spikes placed at the possible values of S_n^*: x_0, x_1, $\ldots$, x_n, where

$$x_j = \frac{j - np}{\sqrt{npq}} \ .$$

- We make the height of the spike at x_j equal to the distribution value $b(n, p, j)$.



- Let ε be the distance between consecutive spikes.

- To change the spike graph so that the area under the curve through the top of the spikes has value 1, we need only multiply the heights of the spikes by $1/\epsilon$.

- We see that

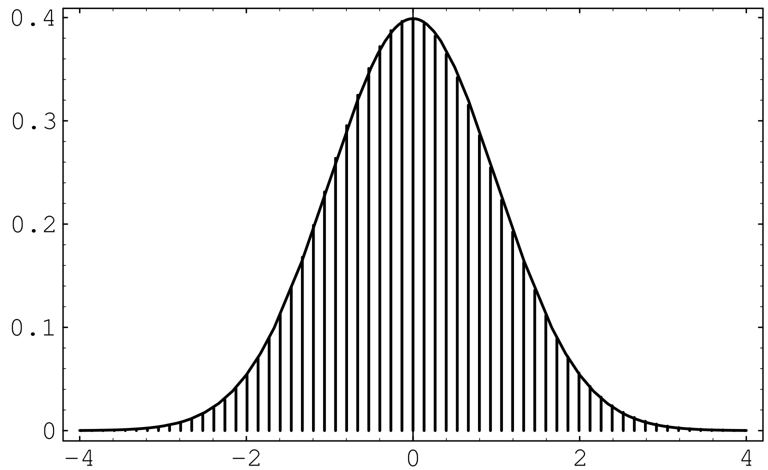
$$\varepsilon = \frac{1}{\sqrt{npq}} \ .$$

- Let us fix a value x on the x-axis and let n be a fixed positive integer.

- Then the point x_j that is closest to x has a subscript j given by the formula
$$j = \langle np + x\sqrt{npq} \rangle \ .$$

- Thus the height of the spike above x_j will be

$$\sqrt{npq}\, b(n, p, j) = \sqrt{npq}\, b(n, p, \langle np + x_j\sqrt{npq} \rangle) \ .$$


0.4
0.3
0.2
0.1
0
−4
−2
0
2
4

Central Limit Theorem for Binomial Distributions

Theorem. *For the binomial distribution* $b(n, p, j)$ *we have*

$$\lim_{n\to\infty} \sqrt{npq}\, b(n, p, \langle np + x\sqrt{npq}\rangle) = \phi(x) \ ,$$

where $\phi(x)$ *is the standard normal density.*

Approximating Binomial Distributions

- To find an approximation for $b(n, p, j)$, we set

$$j = np + x\sqrt{npq}$$

- Solve for x

$$x = \frac{j - np}{\sqrt{npq}} \ .$$

$$\begin{aligned} b(n, p, j) \quad &\approx \quad \frac{\phi(x)}{\sqrt{npq}} \\ &= \quad \frac{1}{\sqrt{npq}}\phi\left(\frac{j - np}{\sqrt{npq}}\right) \ . \end{aligned}$$

Example

- Let us estimate the probability of exactly 55 heads in 100 tosses of a coin.

- For this case $np = 100 \cdot 1/2 = 50$ and $\sqrt{npq} = \sqrt{100 \cdot 1/2 \cdot 1/2} = 5$.

- Thus $x_{55} = (55 - 50)/5 = 1$ and

$$
\begin{aligned}
P(S_{100} = 55) \sim \frac{\phi(1)}{5} &= \frac{1}{5}\left(\frac{1}{\sqrt{2\pi}} e^{-1/2}\right) \\
&= .0484 \ .
\end{aligned}
$$