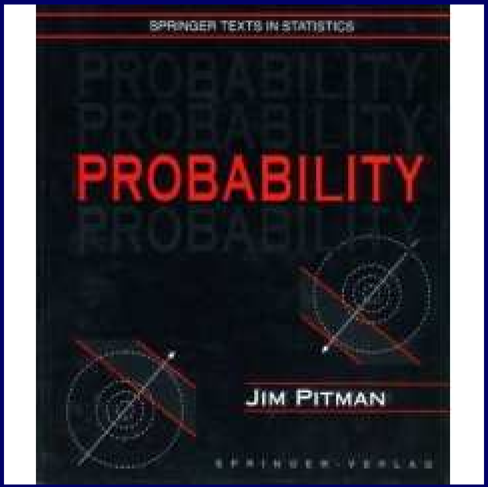
Introduction to probability

Stat 134 FAll 2005

Berkeley



Lectures prepared by:

Elchanan Mossel
Yelena Shvets

Follows Jim Pitman's book:

Probability

Section 4.1

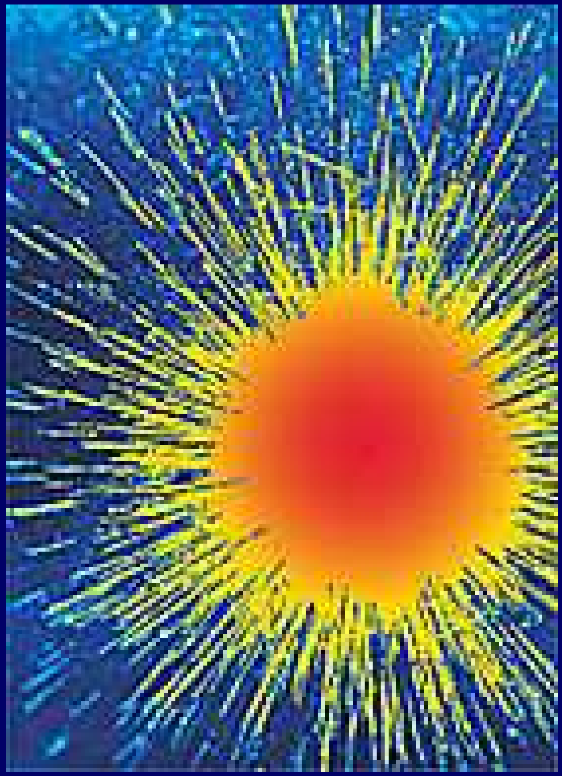
Examples of Continuous Random Variables

<u>Example 1:</u>

X -- The distance traveled by a golf-ball hit by Tiger Woods is a continuous random variable.

PointedMagazine.com

Examples of Continuous Random Variables



Example 2:

Y – The total radioactive emission per time interval is a continuous random variable.

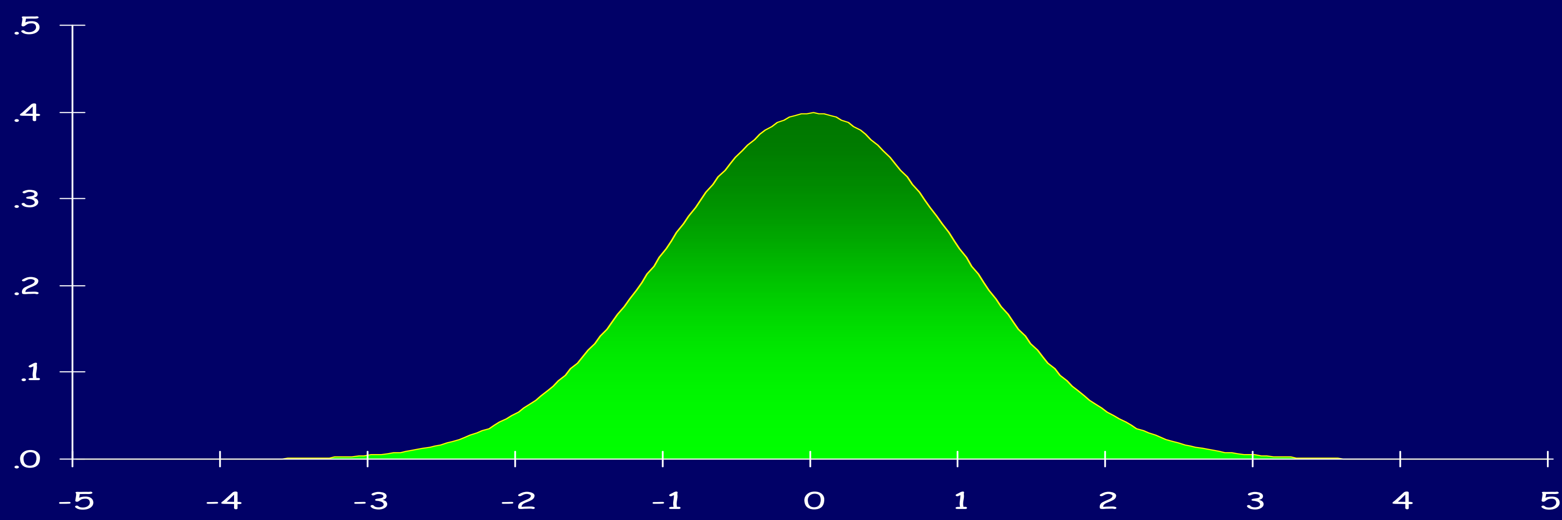
BANG! Colored images of the radioactive emission of a-particles from radium.
C. POWELL, P. FOWLER & D. PERKINS/SCIENCE PHOTO LIBRARY

Examples of Continuous Random variables

Example 3:

$Z \sim N(0,1)$ –

Z approximates (Bin(10000,1/2)-5000)/50 random variable.



How can we specify distributions?

What is

$P\{X_{tiger}=100.1 \text{ ft}\} = ??$

$P\{X_{tiger}=100.0001 \text{ ft}\}=??$

Photo by Allen Eyestone,
www.palmbeachpost.com

Continuous Distributions

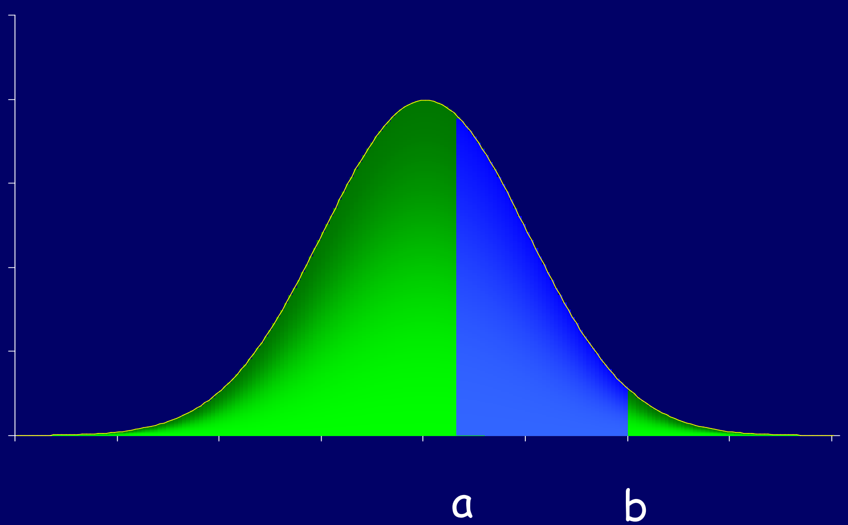
- Continuous distributions are determined by a probability density f.
- Probabilities are defined by areas under the graph of f.
- Example: For the golf hit we will have:

$$P(99.9 < X < 100.1) = \int_{99.9}^{100.1} f_{tiger}(x)dx$$

where f_{tiger} is the probability density of X.

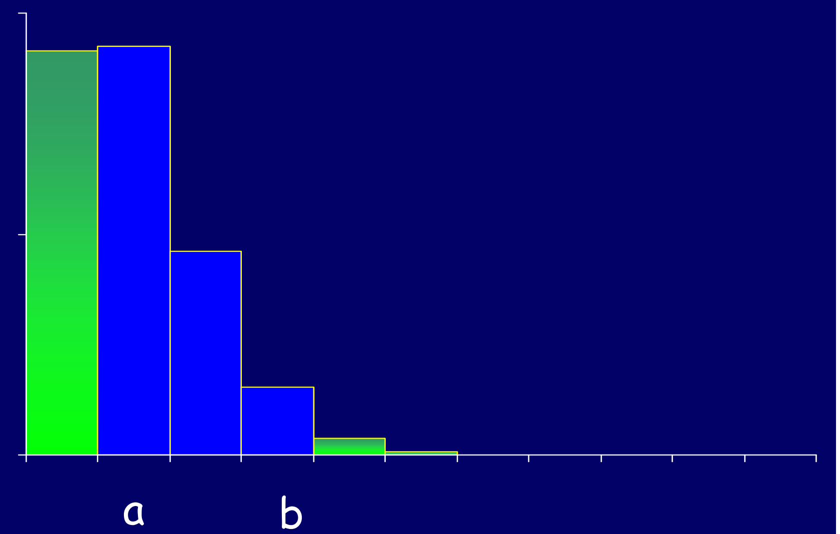
Interval Probability

Continuous



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

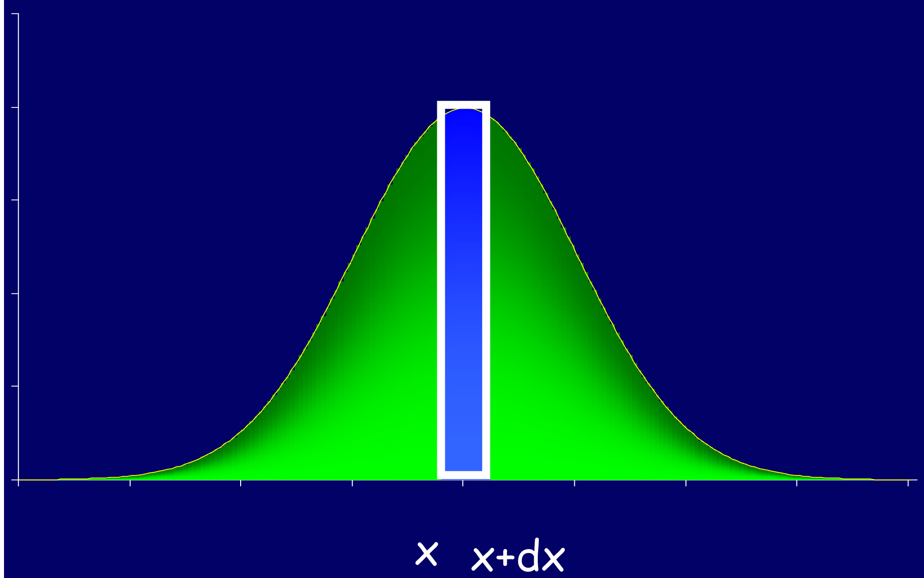
Discrete



$P(a \leq X \leq b) = \sum_{a \leq x \leq b} P(X=x)$

Infinitesimal & Point Probability

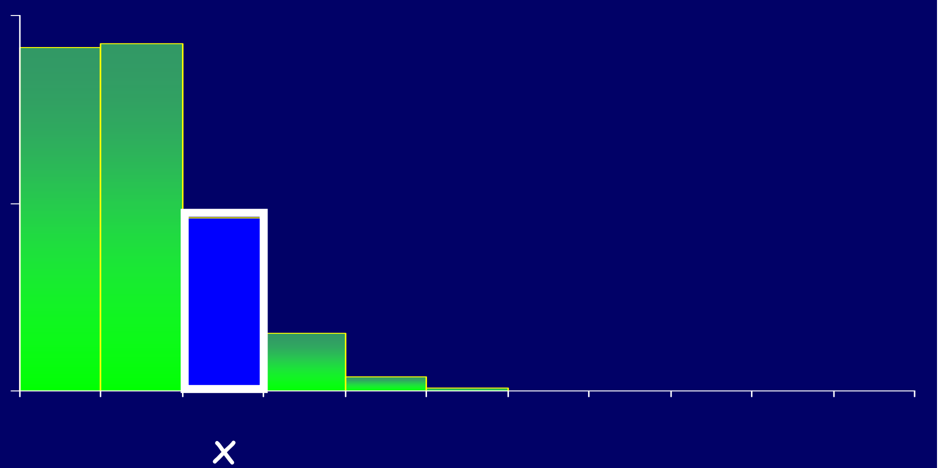
Continuous



$P(x < X \leq x+dx) \approx f(x)\,dx$

(when f is continuous)

Discrete



$P(X=x)=P(x)$

Constraints

Continuous | Discrete

- Non-negative:

Continuous: $f(x) \geq 0$

Discrete: $P(x) \geq 0$

- Integrates to 1:

Continuous: $$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Discrete: $$\sum_{all\ x} P(x) = 1$$

Expectations

Continuous	Discrete

- Expectation of a function g(X):

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)\, dx$$

$$E(g(X)) = \sum_{all\, x} g(x) P(x)$$

- Variance and SD:

$$Var(X) = E[(X - E(X))^2]$$

$$SD(X) = \sqrt{Var(X)}$$

Independence

Continuous | Discrete

- Random Variables X and Y are independent

Continuous

- For all intervals A and B:

$P[X \in A, Y \in B] =$

$P[X \in A]\, P[Y \in B]$.

- This is equivalent to saying that for all sets A and B:

$P[X \in A, Y \in B] =$

$P[X \in A]\, P[Y \in B]$.

Discrete

- For all x and y:

$P[X = x, Y = y] =$

$P[X=x]\, P[Y=y]$.

- This is equivalent to saying that for all sets A and B:

$P[X \in A, Y \in B] =$

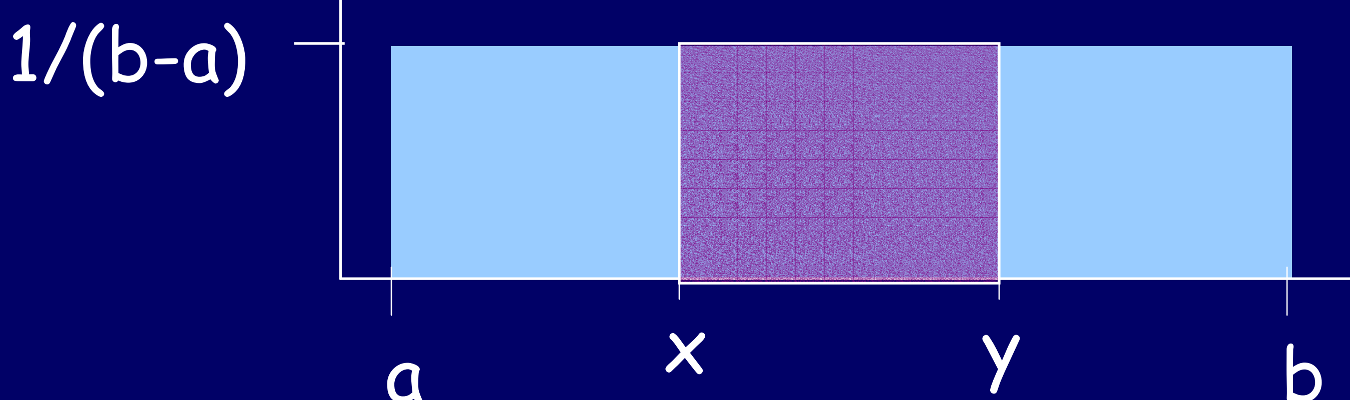
$P[X \in A]\, P[Y \in B]$.

Uniform Distribution

- A random variable X has a <u>uniform distribution on the interval (a,b)</u>,

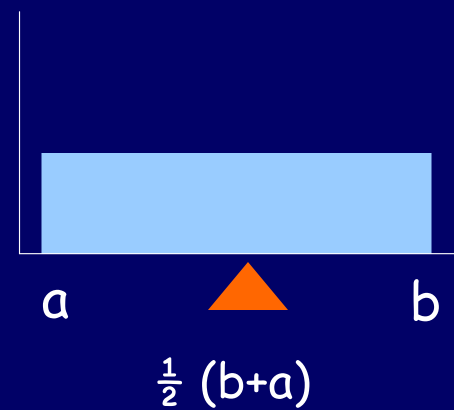
if X has a density f(x) which is

- constant on (a,b),
- zero outside the interval (a,b) and
- $\int_a^b f(x)\,dx = 1$.
- $a \le x < y \le b$: $P(x \le X \le y) = (y-x)/(b-a)$



Expectation and Variance of Uniform Distribution

- $E(X) = \int_a^b x/(b-a)\,dx = \frac{1}{2}(b+a)$



- $E(X^2) = \int_a^b x^2/(b-a)\,dx = 1/3\,(b^2+ba+a^2)$

$Var(X) = E(X^2) - E(X)^2 = 1/3\,(b^2+ba+a^2) - \frac{1}{4}(b^2+2ba+a^2)$

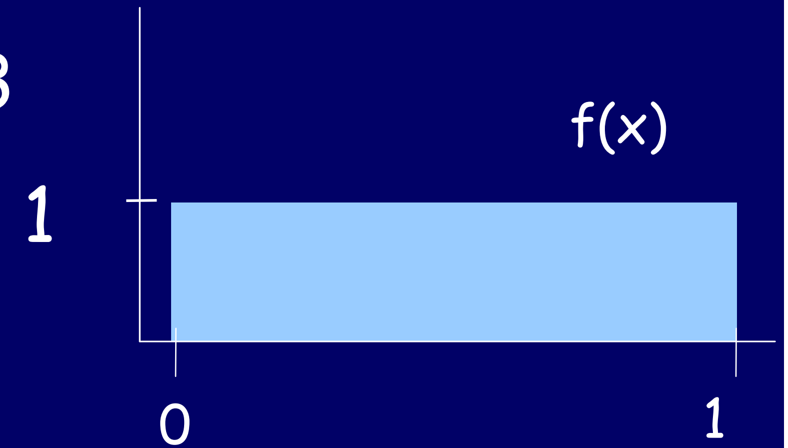
$= \quad 1/12\,(b-a)^2$

Uniform (0,1) Distribution:

- If X is uniform(a,b) then $U = (X-a)/(b-a)$ is uniform(0,1). So:

- $E(U) = \frac{1}{2}$, $E(U^2) = \int_0^1 x^2\,dx = 1/3$

$Var(U) = 1/3 - (\frac{1}{2})^2 = 1/12$



Using $X = U(b-a) + a$:

$E(X) = E[U(b-a) + a] = (b-a)\,E[U] + a$

$= (b-a)/2 + a = (b+a)/2$

$Var(X) = Var[U(b-a) + a] = (b-a)^2 Var[U] = (b-a)^2/12$

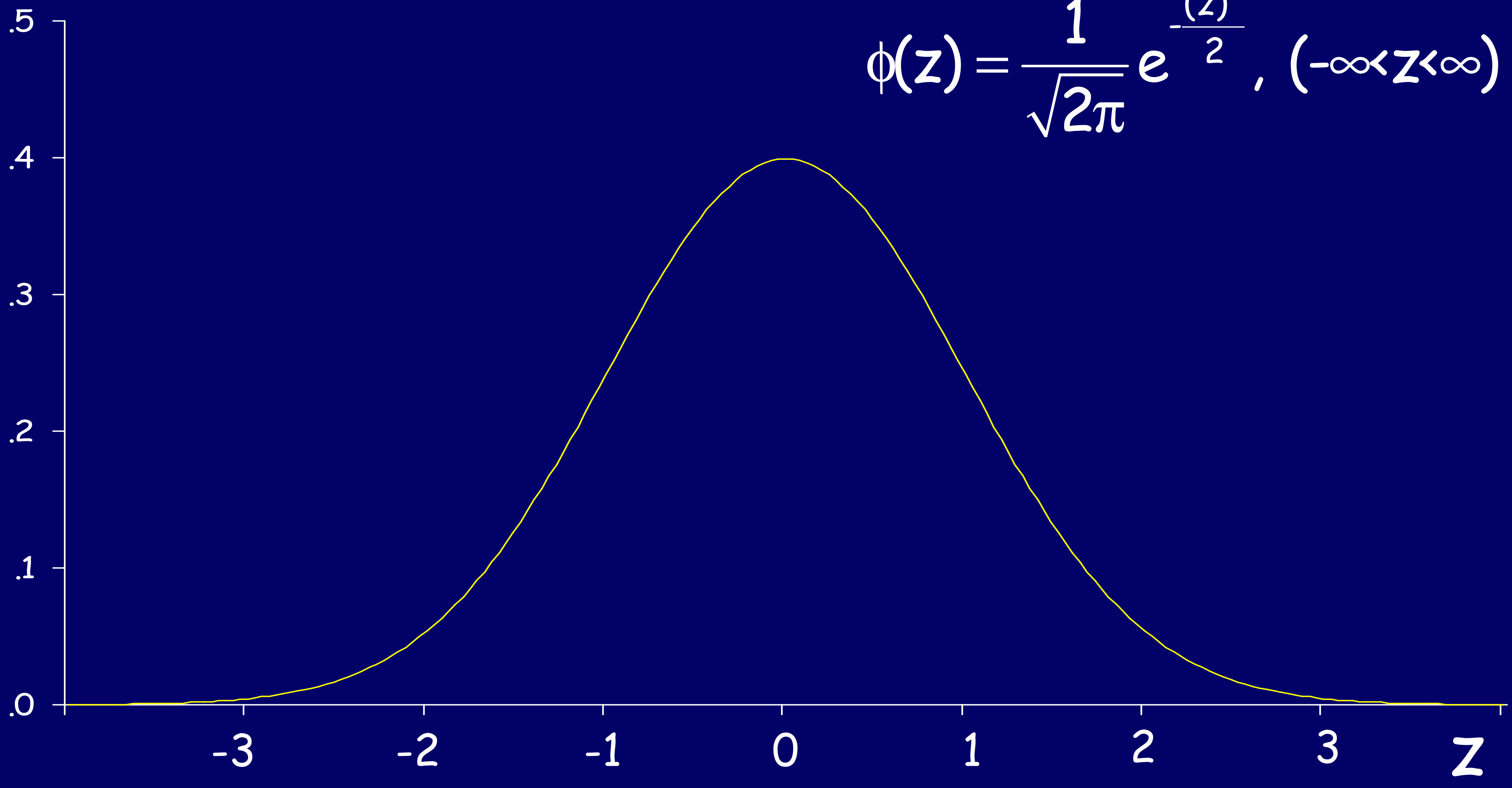
The Standard Normal Distribution

Def: A continuous random variable Z has a standard normal distribution if Z has a probability density of the following form:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}}, \ (-\infty < z < \infty)$$

The Standard Normal Distribution

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}}, \ (-\infty < z < \infty)$$



Standard Normal Integrals

$$\int_{-\infty}^{+\infty} \phi(z)dz = 1$$

$$E(Z) = \int_{-\infty}^{+\infty} z\phi(z)dz = 0$$

$$Var(Z) = \int_{-\infty}^{+\infty} z^2\phi(z)dz = 1$$

Standard Normal Cumulative Distribution Function:

$$\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$$

$P(a \leq Z \leq b) \quad = \quad \Phi(b) - \Phi(a);$

The value of Φ are tabulated in the standard normal table.

The Normal Distribution

Def: If Z has a standard normal distribution and μ and $\sigma > 0$ are constants then X = σ Z + μ has a Normal(μ, σ^2) distribution

Claim: X has the following density:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ (-\infty < x < \infty)$$

We'll see a proof of this claim later

The Normal Distribution

- Suppose $X = \sigma Z + \mu$ has a normal(μ, σ^2) distribution,

Then:

$P(c \le X \le d) = P(c \le \sigma Z + \mu \le d)$

$= P((c-\mu)/ \sigma \le Z \le (d-\mu)/ \sigma)$

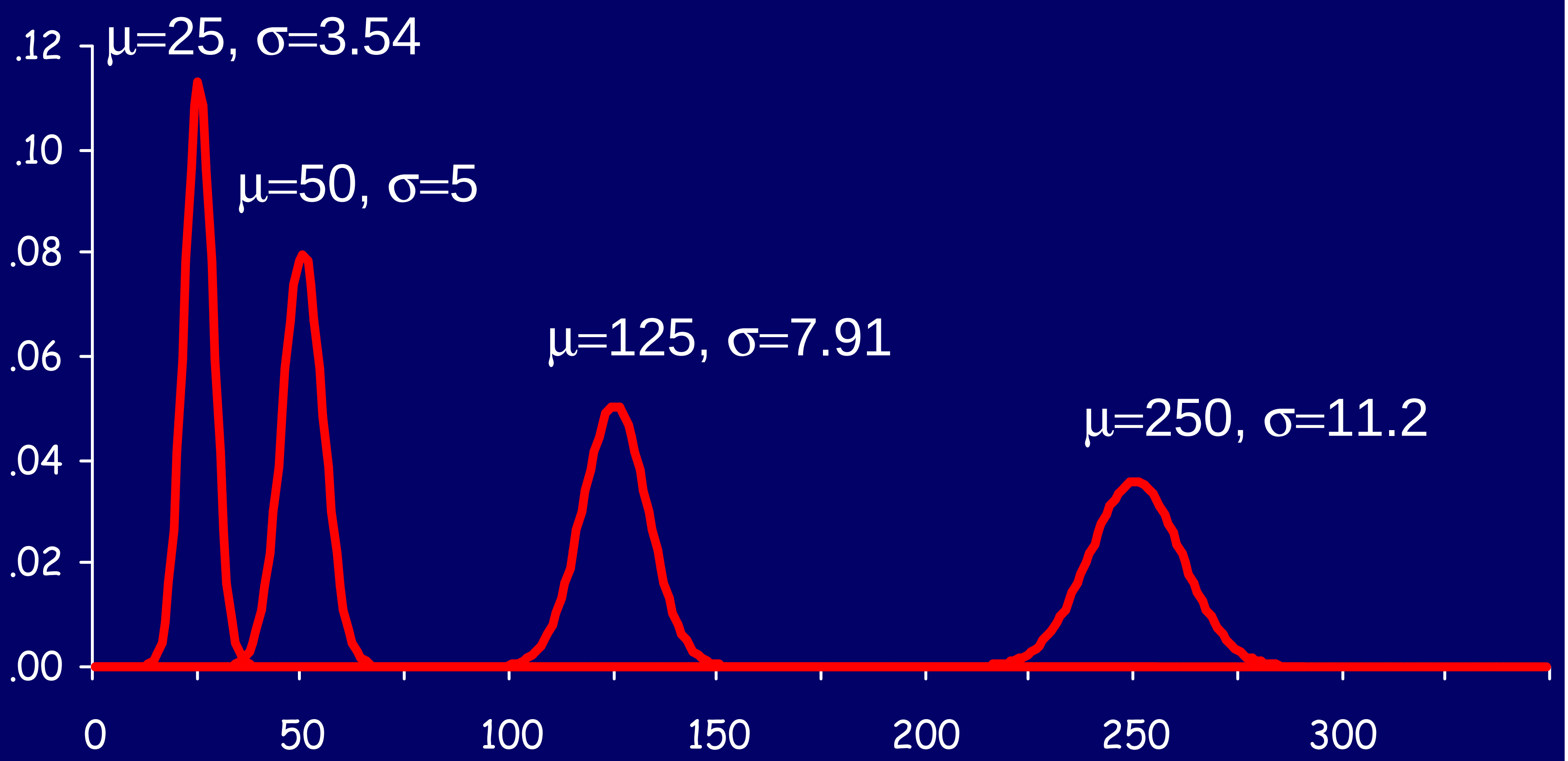
$= \Phi((c-\mu)/\sigma) - \Phi((d-\mu)/\sigma),$

And:

$$E(X) = E(\sigma Z + \mu) = \sigma E(Z) + \mu = \mu$$

$$Var(X) = Var(\sigma Z + \mu) = \sigma^2 E(Z) = \sigma^2$$

Normal(μ,σ^2);



Example: Radial Distance

• A dart is thrown at a circular target of radius 1 by a novice dart thrower. The point of contact is uniformly distributed over the entire target.

- Let R be the distance of the hit from the center. Find the probability density of R.

- Find $P(a \leq R \leq b)$.

- Find the mean and variance of R.

- Suppose 100 different novices each throw a dart. What's the probability that their mean distance from the center is at least 0.7.

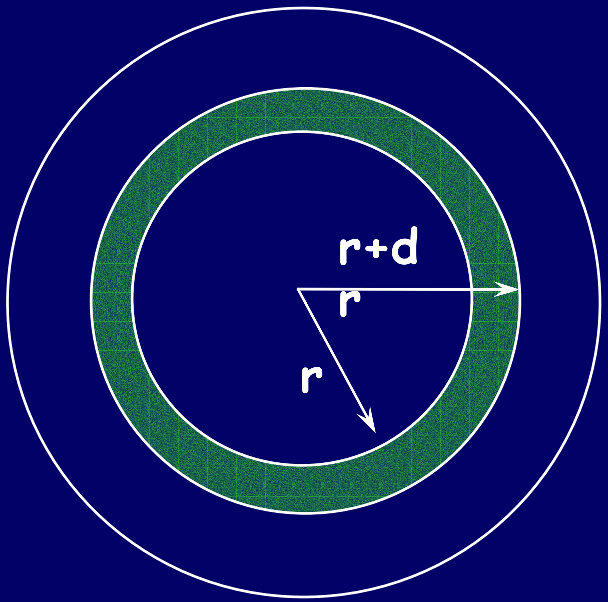
Radial Distance

- Let R be the distance of the hit from the center. Find the probability density of R.

P(R $\in$ (r,r+dr)) = Area of annulus/ total area

= $(\pi(r+dr)^2 - \pi r^2)/\pi$

= 2rdr



$$f(r) = \begin{cases} 2r, & \text{if } 0 \le r \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Radial Distance

With $$f(r) = \begin{cases} 2r, & \text{if } 0 \le r \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

- Find P(a ≤ R ≤ b).

$$P(a \le R \le b) = \int_a^b 2r\,dr = b^2 - a^2$$

- Find the mean and variance of R.

$$E(R) = \int_0^1 2r^2\,dr = \frac{2}{3}$$

$$Var(R) = E(R^2) - E(R)^2 = \int_0^1 2r^3\,dr - E(R)^2$$

$$= \frac{2}{4} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

Radial Distance

- Suppose 100 different novices each throw a dart. What's the probability that their mean distance from the center is at least 0.7.

- Let R_i = distance of the i'th hit from the center, then R_i's i.i.d. with $E(R_i)=2/3$ and $Var(R_i)=1/18$.

- The average $A_{100} = (R_1 + R_2 + ... + R_{100})/100$ is approximately normal with

$$E(A_{100}) = E(R) = \frac{2}{3}$$

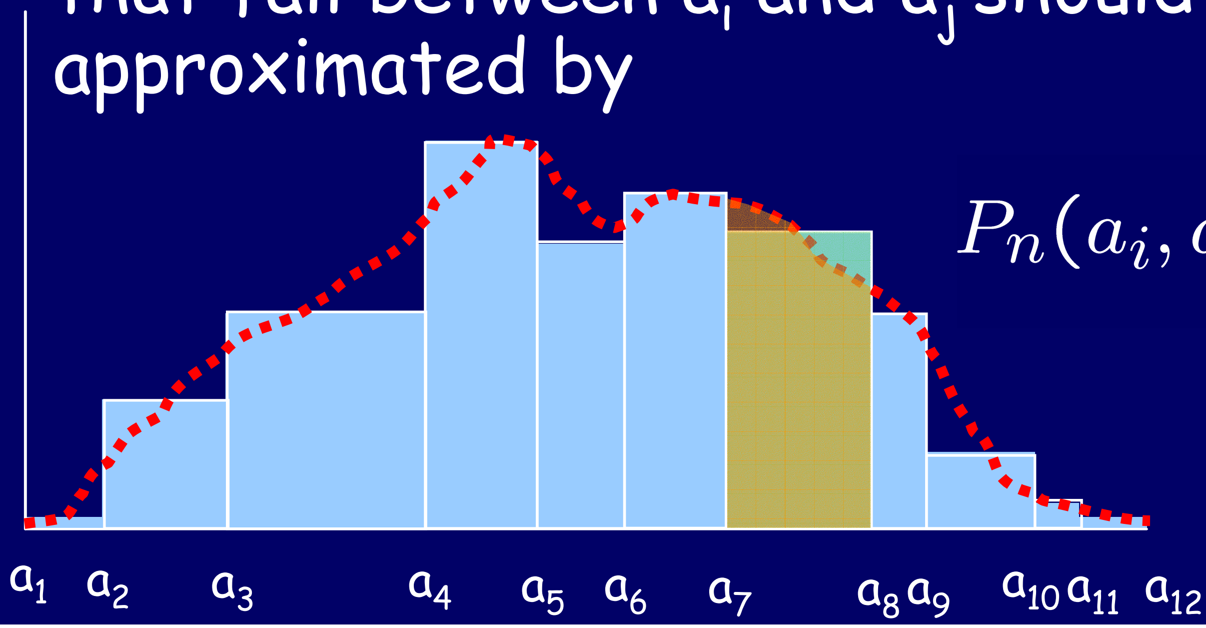
$$SD(A_{100}) = \frac{SD(R)}{\sqrt{100}} = \frac{1}{\sqrt{1800}} \approx 0.0236$$

So:

$$P(A_{100} \geq 0.7) = 1 - \Phi\left(\frac{0.7 - 0.667}{0.0236}\right) = 1 - \Phi(1.40) = 8.7\%$$

Fitting Discrete distributions by Continuous distributions

- Suppose $(x_1, x_2, \ldots, x_n)$ are sampled from a continuous distribution defined by a density f(x).
- We expect that the empirical histogram of the data will be close to the density function.
- In particular, the proportion P_n of observations that fall between a_i and a_j should be well approximated by



$$P_n(a_i, a_j) \approx \int_{a_i}^{a_j} f(x)\, dx.$$

Fitting Discrete distributions by Continuous distributions

- More formally, letting

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

- We expect that:

$$\frac{1}{n}\sum_{i=1}^{n} I_{(a_i,a_j)}(x_i) = P_n(a_i,a_j) \approx \int_{a_i}^{a_j} f(x)\,dx = \int_{-\infty}^{\infty} I_{(a_i,a_j)}(x) f(x)\,dx.$$

- And more generally for functions g we expect:

$$\frac{1}{n}\sum_{i=1}^{n} g(x_i) \approx \int_{-\infty}^{\infty} g(x) f(x)\,dx.$$

Fitting Discrete distributions by Continuous distributions

<u>Claim:</u> if $(X_1, X_2, \ldots, X_n)$ is a sequence of independent random variables each with density f(x) then:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \int_{-\infty}^{\infty} g(x)f(x)\,dx\right| > \epsilon\right) \leq \frac{Var[g(X)]}{n\epsilon^2}.$$

The claim follows from Chebyshev inequality.

Monte Carlo Method

In the Mote-Carlo method the approximation:

$\int g(x)f(x)\,dx \approx 1/n \sum_i g(x_i)$ is used in order to approximate difficult integrals.

Example: $\int_0^1 e^{-\cos(x^3)}\,dx$

take the density to be Uniform(0,1) and

$g(x) = e^{-\cos(x^3)}$.