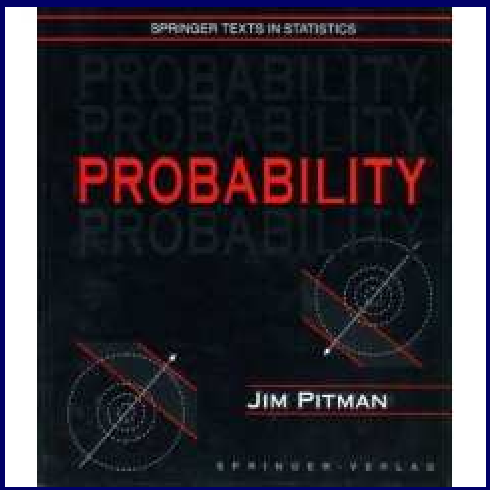
Introduction to probability

Stat 134 FAll 2005

Berkeley



Lectures prepared by:

Elchanan Mossel

Yelena Shvets

Follows Jim Pitman's book:

Probability

Section 4.5

Cumulative Distribution Function

- Definition: For a random variable X, the function

$$F(x) = P(X \leq x),$$

is called the cumulative distribution function (cdf).

- A distribution is called continuous when the cdf is continuous.

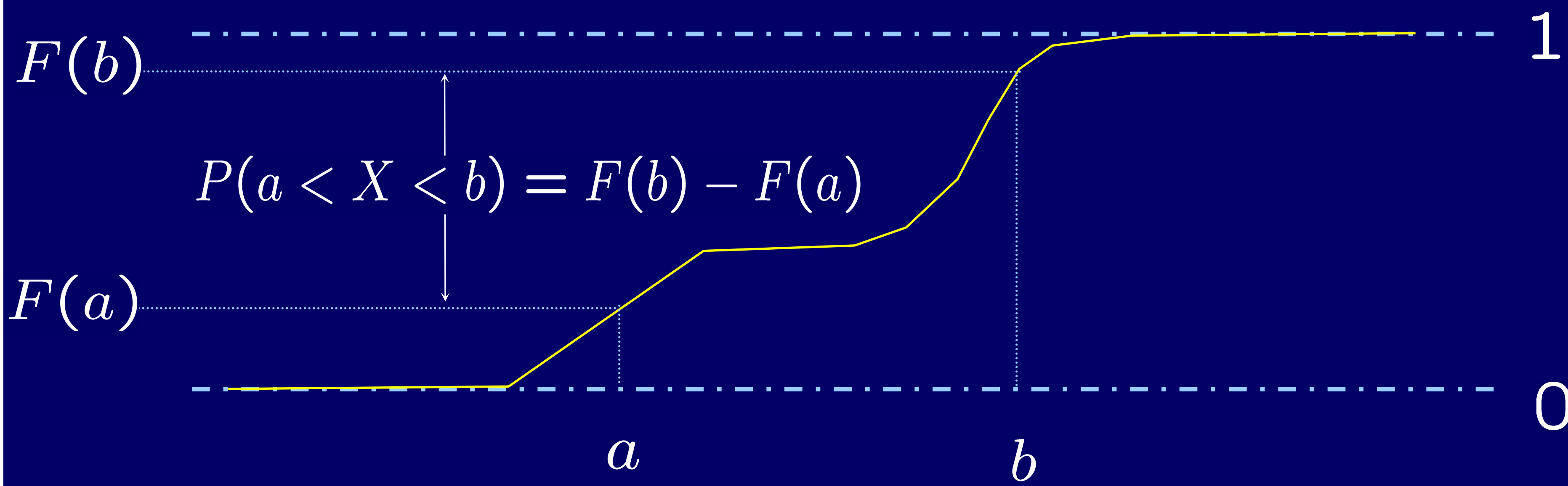
Properties of the CDF.

Claim: The cdf determines the probability of lying inside intervals: $P(a < X \leq b) = F(b) - F(a)$.

- $P\{(a\ ,\ b]\} = F(b) - F(a)$.
- $P\{(a\ ,\ b)\} = \lim_{x \to b^-} F(x) - F(a)$.
- $P\{(a,b) \cap (c,d)\} = P\{(c,b)\}$, for $a<c<b<d$.
- $P\{(a,b) \cup (c,d)\} = P\{(a,d)\}$, for $a<c<b<d$.
- $P\{b\} = P\{(a,b]\} - P\{(a,b)\} = F(b) - \lim_{x \to b^-} F(x)$.

If F is continuous, $P(a,b) = F(b) - F(a)$ and $F(x) = 0$ for all x.

Cumulative Distribution Function

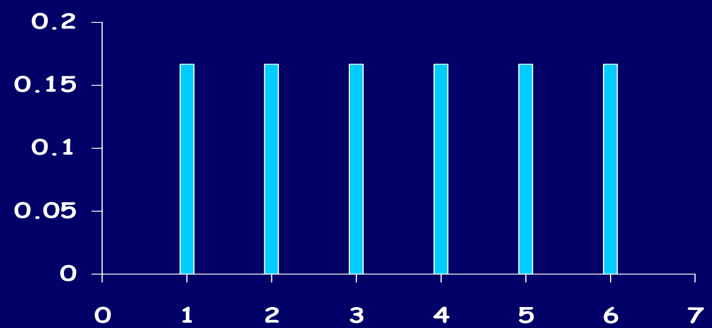


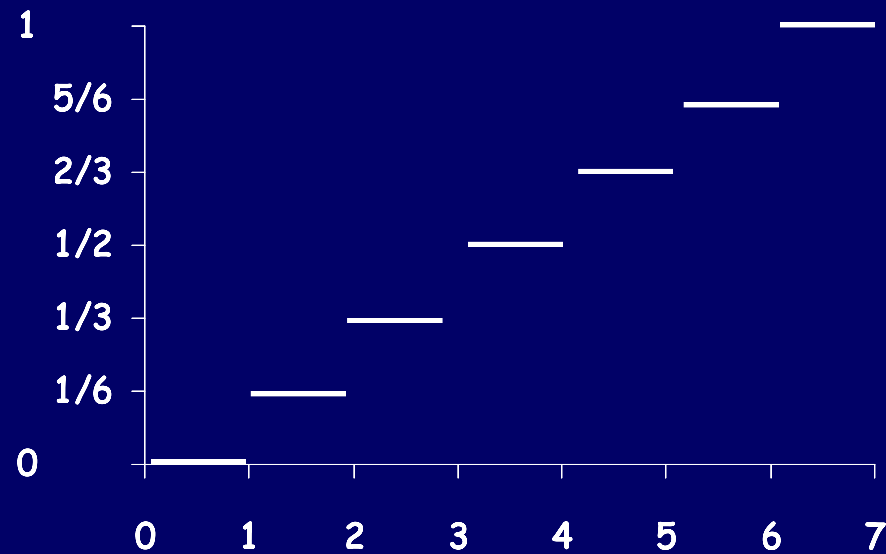
Discrete Distributions

Claim: If X is a discrete variable, then

$F(x) = \sum_{y \leq x} P(X=y)$.

- Example: Let X be the numbers on a fair die.





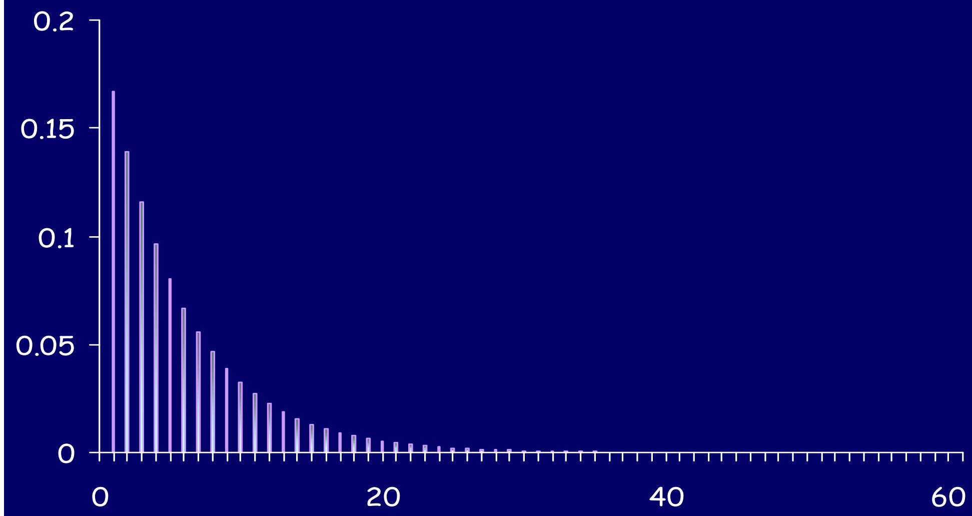
$P\{x\} = F(x) - F(x^-)$: $P\{X=6\} = 1 - 5/6 = 1/6$.

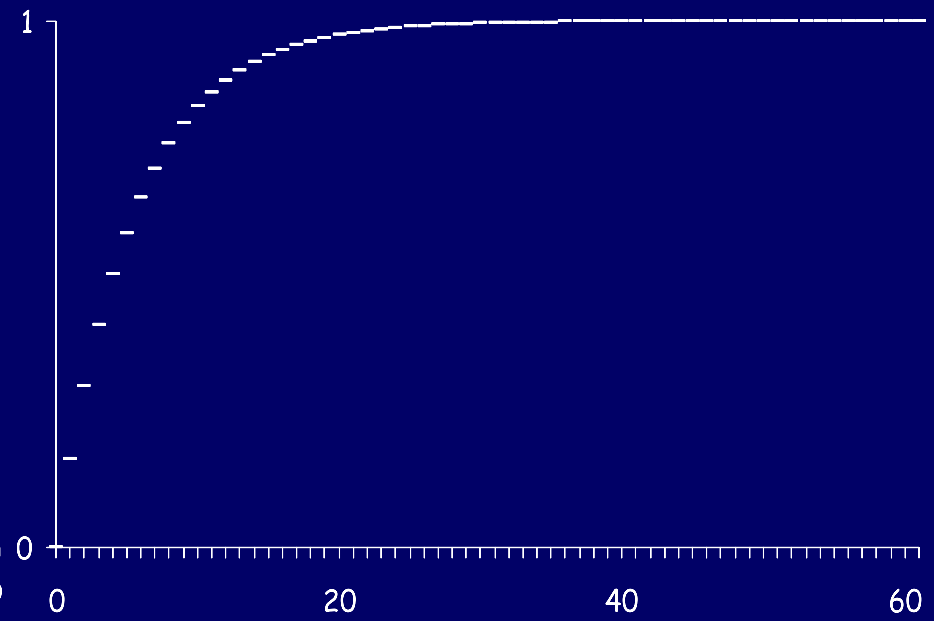
Discrete Distributions

• Example: Let X be the number of rolls before the first six appears when rolling a fair die.

$X \sim$ Geom(1/6), $P(n)=(5/6)^{n-1}\,1/6$ (for integer n)

$F(x) = 1 - (5/6)^{\lfloor x \rfloor}$





For integer x:

$P\{n\} = F(n) - F(n-) = 1 - (5/6)^n - (1 - (5/6)^{n-1}) = (5/6)^{n-1}\,1/6.$

Continuous Distributions with a Density

- Claim: If X is a continuous variable with a density $f_X(x)$ then $F(x) = \int_{-\infty}^{x} f_X(y)dy$.

•Claim: if X has a continuous distribution given by a c.d.f $F(x)$ and

•F is differentiable in all but finitely many points,

•Then the distribution has a density $f_X(x) = F'(x)$.

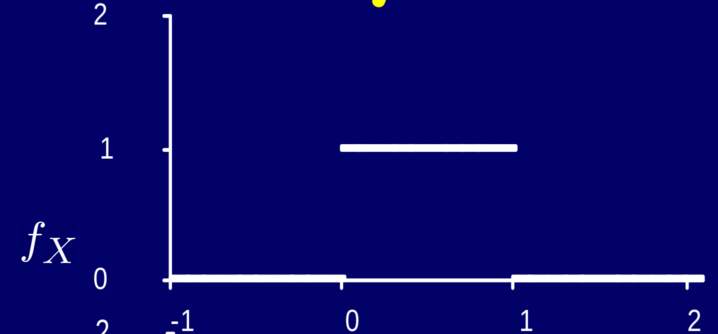
"Pf": We have $F(x) = \int_{-\infty}^{x} f_X(t)\, dt$

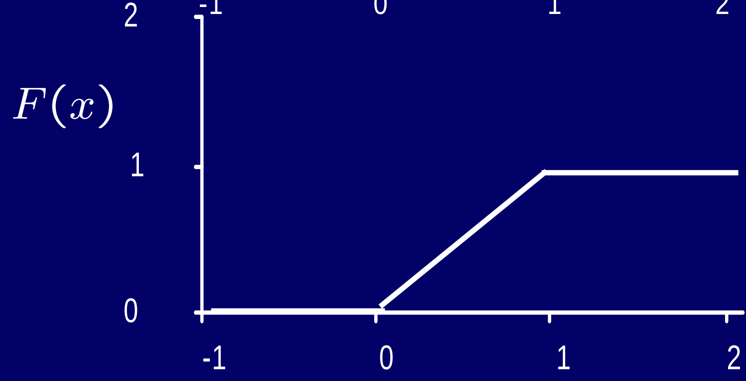
- So: $F'(x) = f_X(x)$ at the points where F is differentiable.

Distributions with a Density

•Example: $U \sim$ Uniform(0,1).

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





Question: Show that $X = 2|U - \frac{1}{2}|$ is a Uniform variable as well.

Solution: $P(X \le x) = P(2|U - \frac{1}{2}| \le x) = P(\frac{1}{2} - x/2 \le U \le \frac{1}{2} + x/2)$
$= F(1/2 + x/2) - F(1/2 - x/2) = 1/2 + x/2 - 1/2 + x/2$
$= x = F(x)$.

So X has the same cdf as U therefore the same distribution.

Question: What is the distribution of $4|\,|U-1/2| - 1/4|$?

Distribution of Min and Max.

Let $X_1, X_2, \ldots, X_n$ be independent random variables.

$X_{max} = \max\{X_1, X_2, \ldots, X_n\}$ & $X_{min} = \min\{X_1, X_2, \ldots, X_n\}$.

It's easy to find the distribution of X_{max} and X_{min} by using the cdf's by using the following observations:

For any x

- $X_{max} \leq x$ iff $X_i \leq x$ for all i.
- $X_{min} > x$ iff $X_i > x$ for all i.

Distribution of Min and Max.

$F_{max}(x) = P(X_{max} \le x) = P(X_1 \le x, X_2 \le x, \ldots X_n \le x)$

$= P(X_1 \le x) P(X_2 \le x) \ldots P(X_n \le x)$, by indep.

$= F_1(x) F_2(x) \ldots F_n(x)$.

$F_{min}(x) = P(X_{min} \le x) = 1 - P(X_{min} > x)$

$= 1 - P(X_1 > x, X_2 > x, \ldots X_n > x)$

$= 1 - (1 - F_1(x))(1 - F_2(x)) \ldots (1 - F_n(x))$.

- $F_{max}(x) = F_1(x) F_2(x) \ldots F_n(x)$.
- $F_{min}(x) = 1 - (1 - F_1(x))(1 - F_2(x)) \ldots (1 - F_n(x))$.

•Example: Suppose X_i's all have exponential distributions with rates λ_i.

Question: Find the distribution of X_{min}.

Solution: For $x < 0$, $F_i(x) = 0$ and for $x \ge 0$, $F_i = 1 - e^{-\lambda x}$.

So for $x < 0$, $F_{min} = 0$ and for $x \ge 0$, $F_{min} = 1 - e^{-\lambda 1 + \lambda 2 + \ldots + \lambda n}$.

$X_{min} \sim \mathrm{Exp}(\lambda_1 + \lambda_2 + \ldots \lambda_n\}$.

Order Statistics

Let X_1, X_2, … X_n be random variables, We can use the CDF to find the CDF of the k^{th} largest of these X. We call this $X_{(k),}$ or the k^{th} order statistic.

The density for the min and max are special cases for k = 1 and k = n.

If all the X_i are identical and independent, we can find the density of the order statistics

Order Statistics

$P(X_{(k)} \sim (x, x+dx)) =$

P(one of the X's in dx)*

P(k-1 of the X's below x)*

P(n-k of the X's above x)=

$$nf(x) * \binom{n-1}{k-1} (F(x))^{k-1} (1-F(x))^{n-k}$$

P(one of the X's in dx)

P(n-k of the X's above x)

$(P(X > x) = 1 - F(x))$

Number of ways to order X's

P(k-1 of the X's below x)

$(P(X < x) = F(x))$

Order Statistics – for Uniform R.V

When X_i are Unif(0,1) we obtain:

$$f_{(k)}(x) = n * \binom{n-1}{k-1} x^{k-1}(1-x)^{n-k}$$

<u>Def:</u> For r,s > 0 the beta(r,s) distribution is the distribution on (0,1) with density $x^{r-1}(1-x)^{s-1}/B(r,s)$ where $B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\, dx$

<u>Claim:</u> If $(X_i)_{i=1}^n$ are i.i.d Unif(0,1) then $X_{(k)}$ has beta(k,n-k+1) distribution.

Beta and Gamma distributions

Claim: If r=k and s=n-k+1 are integers then:

$$B(r,s) = \frac{1}{n\binom{n-1}{k-1}} = \frac{(k-1)!(n-k)!}{n!} = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

where $\Gamma(r) = (r-1)!$ For integer r's.

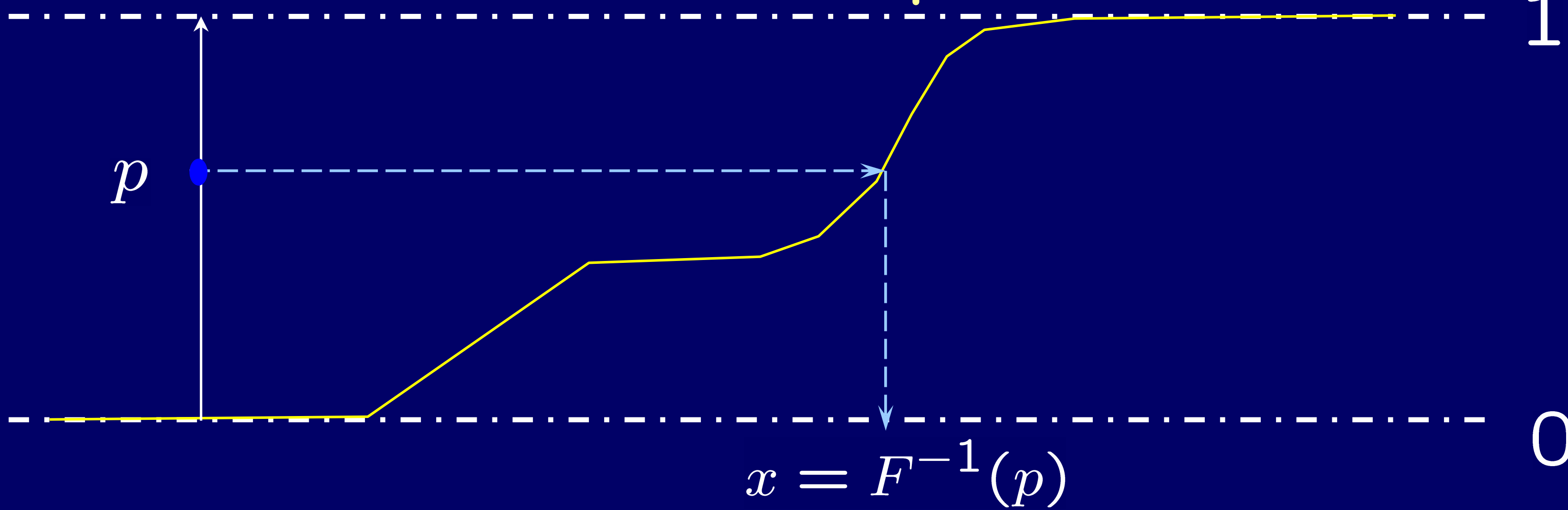
Problem: Calculate E(X) and Var(X) for X with beta(r,s) distribution for integer r and s.

Problem: For the order statistics of n i.i.d. unif(0,1) random variables, calculate E(X) and Var(X).

Percentiles and Inverse CDF.

Definition: The p'th quantile of the distribution of a random variable X is given by the number x such that

$$P(X \leq x) = p.$$



Simulating a RV with a given distribution

Example:

A pseudo random number generator generates a sequence $U_1, U_2, \ldots U_n$ each between 0 and 1, which are i.i.d. Uniform(0,1) random variables.

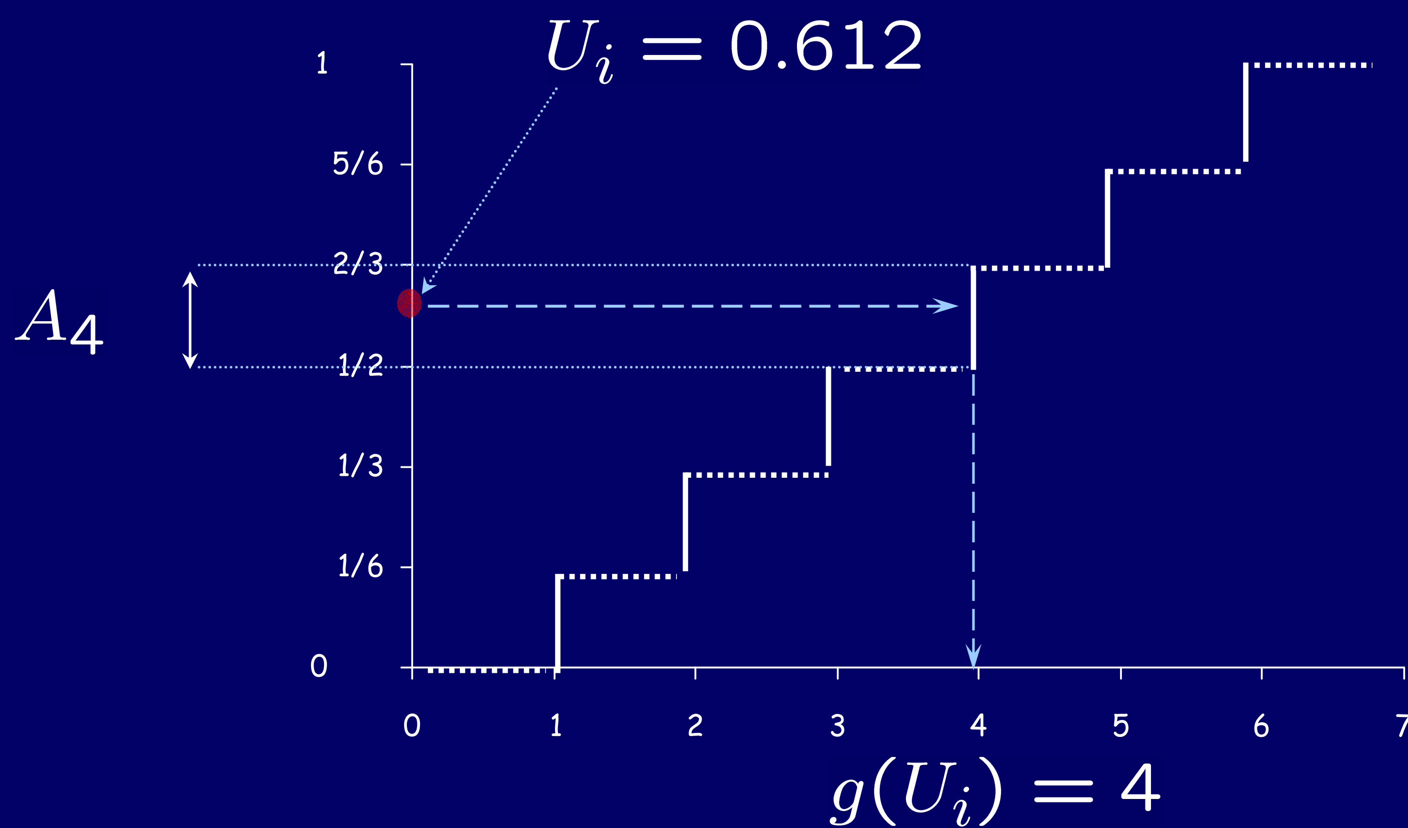
Question: You are working for a company that is developing an on-line casino. How can you use this sequence to simulate n rolls of a fair die?

Solution: You need a new sequence of variables each with a discrete uniform distribution on 1,2,3,4,5,6.

Simple idea -- break up the unit interval into 6 intervals $A_k = ((k-1)/6, k/6]$, each of length 1/6 & let

$$g(u) = k \text{ if } u \in A_k.$$

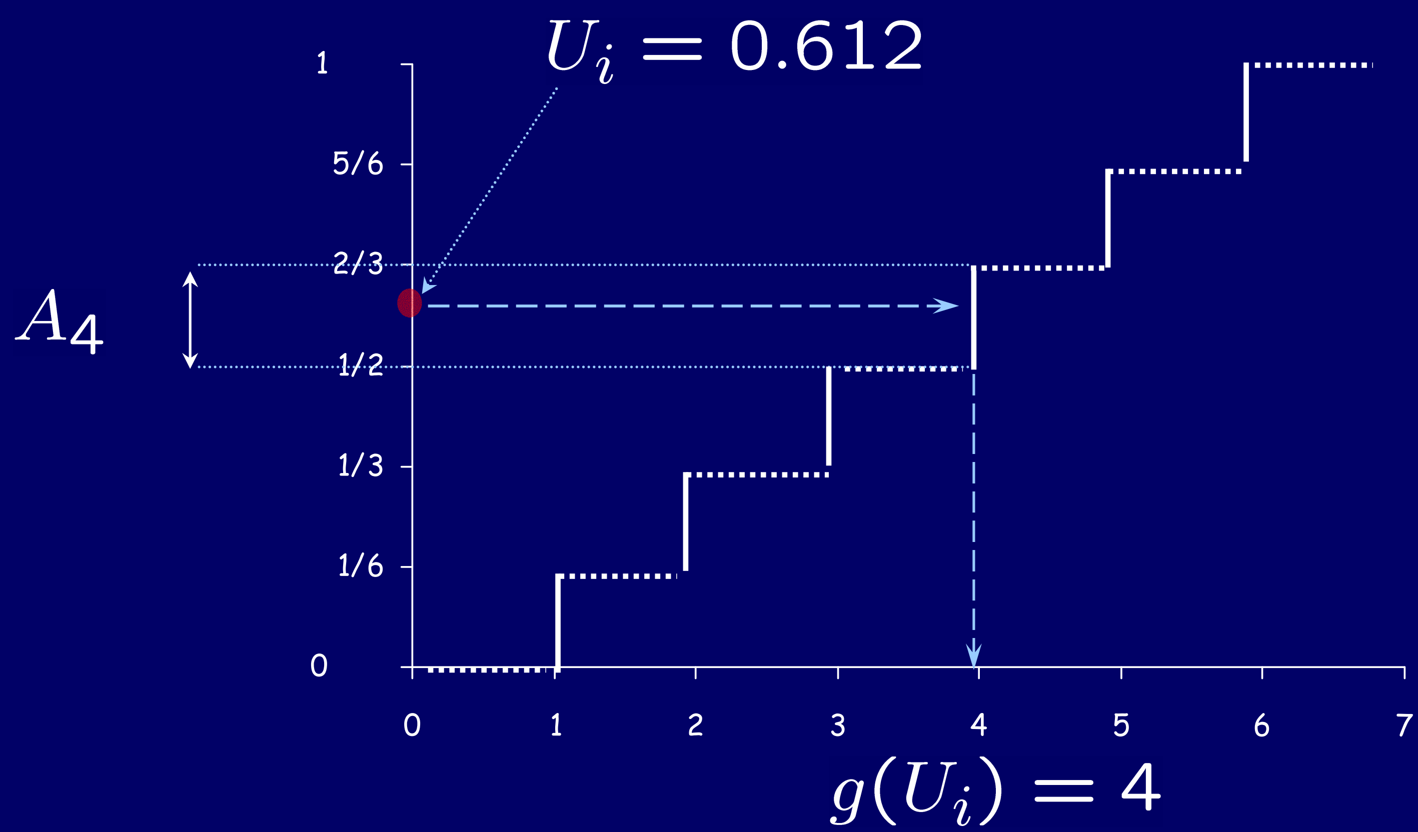
Simulating a RV with a given distribution



Then $X_i = g(U_i)$ each has the desired distribution:

$$P(X_i = k) = P(g(U_i)=k) = P(A_k) = 1/6.$$

Simulating a RV with a given distribution



Note that for i=1,2,3,4,5,6 it holds that:

g(F(i)) = g(i/6)=i.

Simulating a RV with a given distribution

For a general discrete random variable with $P(k) = p_k$, $k=1,2,3\ldots$ we can use the same idea, except each interval A_k is now of length p_k :

$$A_k = (p_1+p_2+\ldots+p_{k-1}\ ,\ \ p_1+p_2+\ldots+p_{k-1}+p_k].$$

The function $g(u)$ defined in this way is a kind of an inverse of the cdf:

$$g(F(k)) = k \text{ for all } k=1,2,3,$$

This observation can be used to simulate continuous random variables as well.

Define $F^{-1}(x) = \min_y F(y) = x$.

Inverse cdf Applied to Standard Uniform

Define $F^{-1}(x) = \min\{ y : F(y) = x\}$

Theorem: For any cumulative distribution function F, with inverse function F^{-1}, if U has a Uniform (0,1) distribution, then a random variable $G = F^{-1}(U)$ has F as a cdf.

Proof (imagine first that F is strictly increasing):

$$P(G \le x) = P\{U \le F(x)) = F(x).$$

Simulating a RV with a given distribution

A pseudo random number generator gives a sequence $U_1, U_2, \ldots U_n$ each between 0 and 1, which behave like a sequence of Uniform(0,1) random variables.

Question: You are working for an engineering company that wants to model failure of circuits. You may assume that a circuit has n components whose life-times have $Exp(\lambda_i)$ distribution. How can you generate n variables that would have the desired distributions?

Solution: The cdf's are $F_i(x) = 1 - e^{-\lambda_i x}$.

So we let $G_i = G(U_i)$ where $G(u) = -\log(1-u)/\lambda_i$.