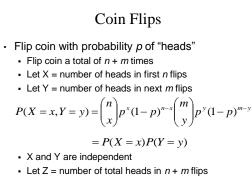
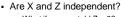
## Independent Discrete Variables

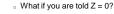
 Two discrete random variables X and Y are called <u>independent</u> if:

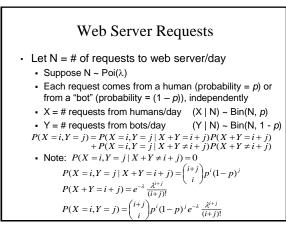
 $p(x, y) = p_x(x)p_y(y)$  for all x, y

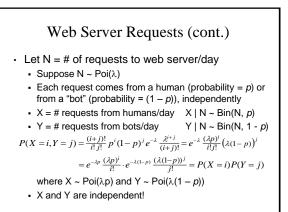
- Intuitively: knowing the value of X tells us nothing about the distribution of Y (and vice versa)
  - If two variables are <u>not</u> independent, they are called <u>dependent</u>
- Similar conceptually to independent *events*, but we are dealing with multiple <u>variables</u>
  - · Keep your events and variables distinct (and clear)!



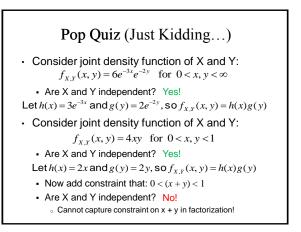


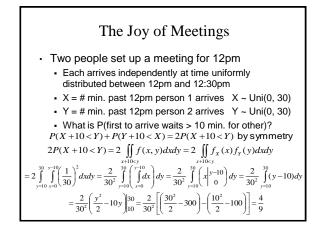






## Independent Continuous Variables • Two continuous random variables X and Y are called <u>independent</u> if: $P(X \le a, Y \le b) = P(X \le a) P(Y \le b)$ for any a, b• Equivalently: $F_{X,Y}(a,b) = F_X(a)F_Y(b)$ for all a,b $f_{X,Y}(a,b) = f_X(a)f_Y(b)$ for all a,b• More generally, joint density factors separately: $f_{X,Y}(x, y) = h(x)g(y)$ where $-\infty < x, y < \infty$

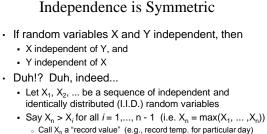




## Independence of Multiple Variables

- *n* random variables X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>n</sub> are called <u>independent</u> if:  $P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$  for all  $x_1, x_2, ..., x_n$
- Analogously, for continuous random variables:

 $P(X_1 \le a_1, X_2 \le a_2, ..., X_n \le a_n) = \prod^n P(X_i \le a_i)$  for all  $a_1, a_2, ..., a_n$ 



- Let event  $A_i = X_i$  is "record value"
  - $_{\circ}$  Is A<sub>n+1</sub> independent of A<sub>n</sub>?
  - Is  $A_{n+1}$  independent of  $A_{n+1}$ ?
  - Easier to answer: Yes!
  - By symmetry,  $P(A_i) = 1/n$  for  $1 \le i \le n$



- From set of *n* elements, choose a subset of size *k* such that all  $\binom{n}{k}$  possibilities are <u>equally</u> likely
  - Only have random (), which simulates X ~ Uni(0, 1)
- Brute force:
  - Generate all subsets of size k
  - Randomly pick one (divide (0, 1) into  $\binom{n}{k}$  intervals)
- · Expensive with regard to time and space
- Bad times!

## (Happily) Choosing a Random Subset

 Good times: int indicator(double p) {

```
\begin{array}{l} \text{ if } (\texttt{random}() < \texttt{p}) \quad \texttt{return 1}; \text{ else return 0}; \\ \\ \text{ subset }\texttt{subset}(\texttt{k}, \texttt{set of size n}) \ \{ \\ \quad \texttt{subset size = 0}; \\ \texttt{I[1] = indicator}((\texttt{double})\texttt{k}/\texttt{n}); \\ \text{ for }(\texttt{i = 1}; \texttt{i < n}; \texttt{i++}) \ \{ \\ \quad \texttt{subset size += I[i]}; \\ \quad \texttt{I[i+1] = indicator}((\texttt{k - subset size})/(\texttt{n - i})); \\ \\ \text{ } \\ \texttt{return (subset containing element[i] iff I[i] == 1}); \\ \\ \\ P(I[\texttt{I}] = \texttt{l}) = \frac{\texttt{k}}{n} \text{ and } P(I[\texttt{i+1}] = \texttt{l} \mid I[\texttt{I}], ..., I[\texttt{i}]) = \frac{\texttt{k} - \sum\limits_{j=1}^{i} I[j]}{n-i}} \text{ where } \texttt{l} < i < n \\ \end{array}
```

