

# Lecture 5

## Statistics

### Distributions

#### Discrete Distributions

what do we need:

$n \rightarrow$  size of  $\Omega$

$\forall i \in \{1, \dots, n\}$

$p_i = P(\{i\})$  - prob. of elementary events

$$E[X]$$

$$Var(X)$$

Examples:

$n=2$

$$\Omega = \{1, 2\} \text{ coin flip} \\ = \{1, 1\}$$

$$P_h \Rightarrow P_1 = 1 - P_2$$

$$h \rightarrow 1 \quad X(h) = 1$$

$$h \rightarrow 0 \quad X(h) = 0$$

$$E[X] = 1 \cdot P(\{1\}) + 0 \cdot P(\{0\})$$

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

$$= P_1 - P_1^2$$

$$= P_1(1 - P_1)$$

0/1 Random variable

$$E[X] = P(X=1)$$

$$E[X^2] = E[X] = P(X=1)$$

$\Leftrightarrow$

$$X^2 = X$$

# Binomial Distribution

Any  $n$

$X \Rightarrow$  # of heads in  $n$  coin flips  
(independent)

Let  $x_i$  be the  $0-1$  RV that models a Bernoulli trial (coin flip)

$$X = \sum_{i=1}^n X_i$$

$$E(X) = \sum_{i=1}^n E(X_i) = np$$

$$E(X^2) = E\left[\sum_{i=1}^n X_i + \sum_{i \neq j} X_i X_j\right]$$

$$= \sum_{i=1}^n \sum_{j=1}^n E(X_i X_j)$$

Underlying prob. dist. is:

$$P(X=k) = P\left\{\underbrace{\{h, \dots, h, t, \dots, t\}}_k \text{ heads} \underbrace{\quad \quad \quad}_{n-k} \text{ tails}\right\}$$

$0 \leq k \leq n$

$$P\left\{\underbrace{\{h, \dots, h, t, \dots, t\}}_k = P\{x_1=h, \dots, x_k=h, x_{k+1}=t, \dots, x_n=t\}\right. \\ \left. = P\{x_1=h\} \dots P\{x_n=t\} = p^k (1-p)^{n-k}\right.$$

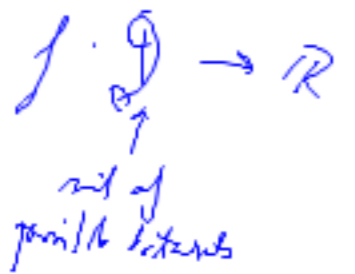
$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E(X) = \sum_{k=0}^n P(X=k) \cdot k \\ = np$$

# Sampling

General Definition: (Problem)

We have some Data  $\mathcal{D}$   
Ex: Dis. random  
Problem: compute  $f(\mathcal{D})$  for some  
function  $f$



$f$  is hard (Adversary) to compute on  $\mathcal{D}$

Sampling solution:

- Pick  $\mathcal{D}'$  a random subset of  $\mathcal{D}$
- Compute  $f(\mathcal{D}')$  and use it to estimate  $f(\mathcal{D})$

Important question:  
what is the probability distribution of random set  $\mathcal{D}'$ ?

Ex:

$P[\mathcal{D}' = \mathcal{D}] = 1$   
 $P[\mathcal{D}' \neq \mathcal{D}] = 0$   
 for some  $\mathcal{D} \subseteq \mathcal{D}$

) not very useful

$\mathcal{D}'$  - first 10 elements of  $\mathcal{D}$

2<sup>o</sup>  $\mathcal{D}' = \{x_1, \dots, x_n\}$  for fixed  $n$  (given)

$P[x_i = y_k] = \frac{1}{|\mathcal{D}|}$   
 $y_k$  is the  $k$ 'th element of  $\mathcal{D}$   
 $\mathcal{D} = \{y_1, \dots, y_{|\mathcal{D}|}\}$

$P[\mathcal{D}'] = P[\{x_1, \dots, x_n\}]$   
 $= P[x_1] \dots P[x_n]$   
 $= \left(\frac{1}{|\mathcal{D}|}\right)^n$

) Sampling with replacement

## Sampling without replacement

Once an element is picked, it is not returned to  $D$  (and is picked again)

$$P(D^c) = P\{x_1\} \cdot P\{x_2 | x_1\} \cdot \dots \cdot P\{x_n | x_1, x_2, \dots, x_{n-1}\}$$

$$P(A, B) = P(A) \cdot P(B|A)$$

$$= \frac{1}{|D|} \cdot \frac{1}{|D|-1} \cdot \dots \cdot \frac{1}{2} \cdot \frac{1}{1}$$
$$= \frac{1}{|D|!}$$

$P(A, B)$  characters are just for  $A$  and let  $B$  be the first  $n$  elements