TDAB01 Probability and Statistics

Maryna Prus
IDA, Linköping University

Lecture 2: Discrete Random Variables

Overview

- **Random variables**
- **Probability distribution**
- **Expected value and variance**
- **Covariance and correlation**
- **Chebyshev's inequality**

Random variables

Definition. A **random variable** X is a function from Ω to $\mathbb{R}$:

$$X = f(\omega)$$

where $\omega \in \Omega$ is an outcome.

- Two types of random variables:
  - **Discrete:** X can take values from a finite (for example $\{0, 1, 2, \ldots, n\}$) or countable (for example $\{0, 1, 2, \ldots\}$) set
  - **Continuous**: X can take all values from an interval, especially $(0, \infty)$ or $\mathbb{R}$
- Random variables are also called **stochastic variables**

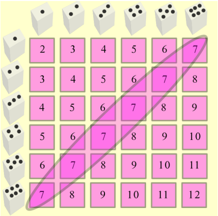
Random variables: Some examples

- Example: Tossing two coins
  - $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$
  - X = number of H

$$
\begin{aligned}
\boldsymbol{P}(X=0) &= \boldsymbol{P}(\{(T,T)\}) = 1/4 \\
\boldsymbol{P}(X=1) &= \boldsymbol{P}(\{(H,T),(T,H)\}) = 1/2 \\
\boldsymbol{P}(X=2) &= \boldsymbol{P}(\{(H,H)\}) = 1/4
\end{aligned}
$$

- Example: Rolling two dice
  - $\Omega = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$
  - X = total number of dots (sum)

$$
\begin{aligned}
\boldsymbol{P}(X=2) &= \boldsymbol{P}(\{(1,1)\}) = 1/36 \\
\boldsymbol{P}(X=3) &= \boldsymbol{P}(\{(1,2),(2,1)\}) = 1/18 \\
\boldsymbol{P}(X=4) &= \boldsymbol{P}(\{(1,3),(2,2),(3,1)\}) = 1/12 \\
&\ldots
\end{aligned}
$$



Probability distribution

Definition. **(Probability) distribution** of X is the collection of all probabilities related to X. **Probability (mass) function** of X:

$$P(x) = \boldsymbol{P}(X = x)$$

for all values of x.

- $0 \le P(x) \le 1$, for all x
- To notation:
  - X - **random variable**; for example, sum of dots on two dice
  - x - **given outcome**; for example, 7 dots
- For **discrete** random variables $P(x)$ is also calles **pmf** (probability mass function)
- Set of all possible values of X is called **support** of distribution of X (or **support** of X)

Distribution function

Definition. **(Cumulative) distribution function** of X is defined as

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

for all values of x.

- Probabilities for all values of x sum up to 1:

$$\sum_{\text{all } x} P(x) = \sum_{\text{all } x} \boldsymbol{P}(X = x) = 1$$

- Distribution function is non-decreasing, always between 0 and 1,

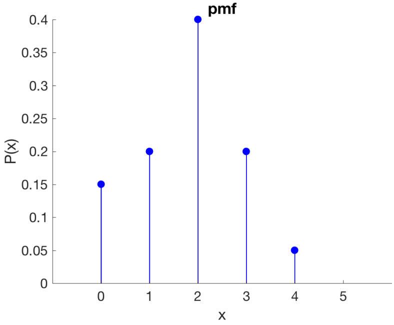
$$\lim_{x \to -\infty} F(x) = 0 \quad \& \quad \lim_{x \to +\infty} F(x) = 1$$

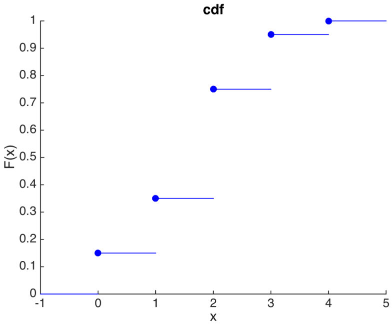
- Distribution function is also called **cdf** (cumulative distribution function)

- $P(a < X \leq b) = F(b) - F(a)$

Probability and distribution function

x	0	1	2	3	4
$P(x)$	0.15	0.20	0.40	0.20	0.05
$F(x)$	0.15	0.35	0.75	0.95	1.00





Joint distribution

- X and Y - random variables $\rightarrow$ (X, Y) - **random vector**
- Distribution of (X, Y) is called **joint distribution** of X and Y
- **Joint probability function**:

$$P(x, y) = \boldsymbol{P}((X, Y) = (x, y)) = \boldsymbol{P}(X = x \cap Y = y) = \boldsymbol{P}(X = x, Y = y)$$

- Probabilities for all values of (x, y) sum up to 1:

$$\sum_{\text{all } x} \sum_{\text{all } y} P(x, y) = 1$$

- Example: X = Spam / Ham (not Spam) and Y = Inbox / Spambox

	Spam	Ham
Inbox	0.02	0.88
Spambox	0.09	0.01

- Joint distribution: "What is the probability of getting a ham email **and** that it ends up in the spam folder?"

Joint distribution

- Example: X return on stock X, and Y return on stock Y

		Stock Y		
		Low	Medium	High
	Low	0.05	0.05	0.15
Stock X	Medium	0.10	0.30	0.20
	High	0.05	0.05	0.05

- Stock portfolio: 50% of stock X and 50% of stock Y
- Joint distribution: "What is the probability of medium return on stock portfolio?"

Marginal distribution

- Distribution of (X, Y) - joint distribution of X and Y
- Distribution of X - **marginal distribution** (of X)
- Distribution of Y - marginal distribution (of Y)
- Marginal distribution can be obtained from joint distribution:

$$P_X(x) = \sum_{\text{all } y} P(x, y)$$

$$P_Y(y) = \sum_{\text{all } x} P(x, y)$$

- Example: X =Spam/Ham och Y =Inbox/Spambox

	Spam	Ham	
Inbox	0.02	0.88	**0.9**
Spambox	0.09	0.01	**0.1**
	0.11	**0.89**	

- Marginal distribution: "What is the probability of receiving a spam email (regardless of where it ends up) ?"

Marginal distribution

- Exempel: X = return on stock X, and Y = return on stock Y

		Stock Y			
		Low	Medium	High	
Stock X	Low	0.05	0.05	0.15	0.25
	Medium	0.10	0.30	0.20	0.6
	High	0.05	0.05	0.05	0.15
		0.20	0.40	0.40	

- Marginal distribution: "What is the probability of high return on stock Y (regardless of return on stock X) ?"

Independence

Definition. Random variables X and Y are **independent** if

$$P(x, y) = P_X(x) \cdot P_Y(y)$$

for **all** values x and y

- Example: X = Spam/Ham and Y = Inbox/Spambox

	Spam	Ham	
Inbox	0.02	0.88	0.9
Spambox	0.09	0.01	0.1
	0.11	0.89	

- X and Y are not independent:

$$P(\text{inbox}) \cdot P(\text{ham}) = 0.9 \cdot 0.89 = 0.801 \neq 0.88 = P(\text{inbox}, \text{ham})$$

- Alternatively

$$P(\text{inbox}|\text{ham}) = \frac{P(\text{inbox}, \text{ham})}{P(\text{ham})} = \frac{0.88}{0.89} = 0.988 > 0.9 = P(\text{inbox})$$

Expected value

Definition. **Expectation** or **expected value** of a random variable X is its mean or average value:

$$\mu = \mathbb{E}(X) = \sum_{\text{all } x} x \cdot P(x).$$

- Example

x	0	1	2	3	4
$P(x)$	0.15	0.20	0.40	0.20	0.05

- Expected value:

$$\mathbb{E}(X) = 0 \cdot 0.15 + 1 \cdot 0.20 + 2 \cdot 0.40 + 3 \cdot 0.20 + 4 \cdot 0.05 = 1.8$$

- g - function of X

$$\mathbb{E}(g(X)) = \sum_{\text{all } x} g(x) \cdot P(x)$$

Variance

- Deviations $x - \mathbb{E}(X)$ give some information about spread
- Expected deviation:

$$\mathbb{E}(X - \mu) = \sum_{\text{all } x} P(x) \cdot (x - \mu)$$

- Problem: $\mathbb{E}(X - \mu)$ is zero
- **Variance** - Expected squared deviation from the mean:

$$\sigma^2 = Var(X) = \mathbb{E}\left[(X - \mu)^2\right] = \sum_{\text{all } x} (x - \mu)^2 \cdot P(x)$$

- Alternative formula

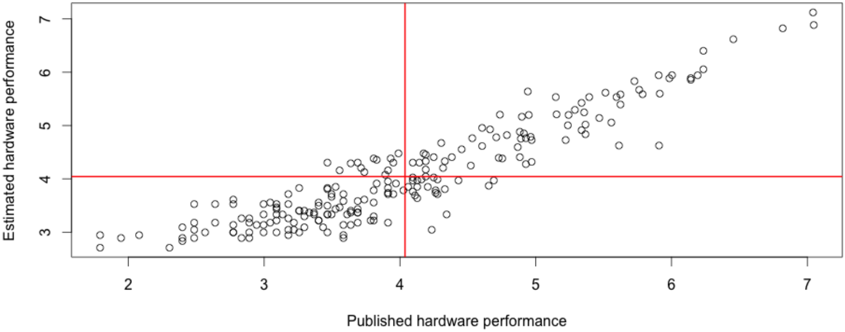
$$Var(X) = \mathbb{E}(X^2) - \mu^2$$

- **Standard deviation**: $\sigma = Std(X) = \sqrt{Var(X)}$

Properties of expectation and variance

- $\mathbb{E}(c) = c$, where c is a constant
- $\mathbb{E}(aX + b) = a \cdot \mathbb{E}(X) + b$, for constant a, b
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
- $\mathbb{E}(aX + bY + c) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y) + c$, for constant a, b, c
- $Var(aX + b) = a^2 \cdot Var(X)$
- X and Y independent: $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- X and Y independent: $Var(X + Y) = Var(X) + Var(Y)$

Covariance and correlation



Covariance and correlation

- **Covariance** between X and Y:

$$\begin{aligned}\sigma_{XY} = Cov(X,Y) &= \mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))] \\ &= \sum_{\text{all } x}\sum_{\text{all } y}(x-\mu_x)\cdot(y-\mu_y)\cdot P(x,y)\end{aligned}$$

  where $\mu_x = \mathbb{E}(X)$ and $\mu_y = \mathbb{E}(Y)$
- σ_{XY} illustrates linear dependence between X and Y
- Positive covariance:
  - Larger values of X correspond to larger values of Y; Smaller values of X correspond to smaller values of Y
- Negative covariance:
  - Larger values of X correspond to smaller values of Y; Smaller values of X corresponds to larger values of Y
- **Correlation coefficient** between X and Y

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{Std(X)\cdot Std(Y)}$$

Properties of covariance and correlation

- $Cov(X,Y) = Cov(Y,X)$
- $\sigma^2 = \sigma_{XX} = Cov(X,X)$
- $Var(aX + bY + c) = a^2 \cdot Var(X) + b^2 \cdot Var(Y) + 2 \cdot a \cdot b \cdot Cov(X,Y)$, for constant a, b, c
- $Cov(a \cdot X + b, c \cdot Y + d) = a \cdot c \cdot Cov(X,Y)$, for constant a, b, c
- $-1 \leq \rho \leq 1$
- If $|\rho| = 1$, then Y is a linear function of X
- If X and Y independent, then $Cov(X,Y) = 0$ and $\rho(X,Y) = 0$
- But $Cov(X,Y) = 0$ or $\rho(X,Y) = 0$ **do not necessarily** mean independence of X and Y

Chebyshev's inequality

- For any distribution with expected value μ and variance σ^2 and for all $\varepsilon > 0$ the following inequality holds:

$$\boldsymbol{P}(|X - \mu| > \varepsilon) \leq \left(\frac{\sigma}{\varepsilon}\right)^2$$

\- **Chebyshev's inequality**

- Interpretation: X lie in the interval $[\mu - \varepsilon, \mu + \varepsilon]$ with a probability which is at least $1 - (\sigma/\varepsilon)^2$
- Chebyshev's inequality only requires knowledge of μ and σ^2
  No other properties of the distribution of X
- $\left(\frac{\sigma}{\varepsilon}\right)^2$ is often much larger than the true probability $\boldsymbol{P}(|X - \mu| > \varepsilon)$
- Example: see Example 3.12 in textbook

Thank you for your attention!