Cumulative Distribution Function

PMF $P_x(x_k) = \frac{1}{4}$ PMF discrete random variable

→ CDF can be used for any kind of random variable

- Definition

A CDF of a random variable X is defined as $F_x(x) = P(X \le x)$ for all $x \in R$

EX: I have tossed a coin twice. RV X is number of heads. Find its PMF and CDF.

$R_x = \{0,1,2\}$ $\underbrace{P_x(0) = \frac{1}{4} \quad P_x(1) = \frac{1}{2} \quad P_x(2) = \frac{1}{4}}_{PMF}$

CDF: $F_X(x) = P(X \le x)$

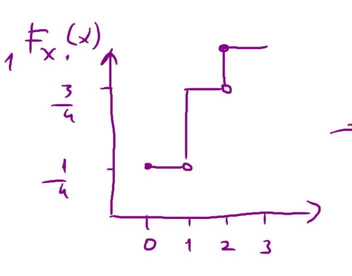
For $\alpha < 0$ $F_x(\alpha) = 0$

$\alpha \ge 2$ $F_x(\alpha) = 1 = P_x(0) + P_x(1) + P_x(2)$

$0 \le \alpha < 1$ $F_x(\alpha) = \frac{1}{4}$

$1 \le \alpha < 2$ $F_x(\alpha) = \frac{3}{4} = P_x(0) + P_x(1)$

$$F_x(x) = \begin{cases} 0 & \alpha < 0 \\ \frac{1}{4} & 0 \le \alpha < 1 \\ \frac{3}{4} & 1 \le \alpha < 2 \\ 1 & 2 \le \alpha \end{cases}$$



→ CDF

Properties of Cumulative Distribution Function

1-) $F_x(-\infty) = 0$ $F_x(\infty) = 1$

2-) CDF is a $\underbrace{\text{non decreasing}}_{\text{azalmayan}}$ function

$\alpha \le \beta \Rightarrow F_x(\alpha) \le F_x(\beta)$

3-) For any $x_k \in R_y$

$F_x(x_k) - F_x(x_k - \epsilon) = P_x(X = x_k)$

epsilon

4-) $P(a < x \le b) = P_x(X \le b) - P_x(X \le a) = F_x(b) - F_x(a)$

EXPECTATION

$a_1, a_2 \dots a_n$ → average $\underbrace{\sum_{i=0}^{n} a_i}_{\text{mean}}$ average

Consider a R.V. X, how we find its average?

Let X be a R.V. with $R_x = \{x_1, x_2, \dots \infty\}$ the expected value of random variable X, EX

$EX = \sum_{x_k \in R_x} x_k \cdot P(X = x_k)$

Expected value

$E[X] = E(X) = \mu_X$

EX:

X	P(x)
2	0.75
3	0.25

EX of $X = \sum x_k \cdot P_x(x_k)$

$= 2 \cdot 0.75 + 3 \cdot 0.25$

$= 2.25 \to EX$

FUNCTIONS OF RANDOM VARIABLES

X is a random variable $R_y = \{g(x) \mid x \in R_x\}$

$Y = g(X)$ → R.V.; function; R.V.

we already know the PMF(X), PMF(Y)

$P_y(y) = P(Y = y) = P(g(x) = y) = \sum_{x: g(x) = y} P_x(x)$

EX: Let X be a discrete R.V. with $P_X(k) = \frac{1}{5}$ for $k = -1, 0, 1, 2, 3$

Let $Y = 2|X|$ find range and pmf of Y.

$R_x = \{-1, 0, 1, 2, 3\}$ $P_x(0) = P_x(1) \dots P_x(3)$ $R_y = \{|2 \cdot -1|, |2 \cdot 0|, |2 \cdot 1|, |2 \cdot 2|, |2 \cdot 3|\}$ → $\{2, 0, 4, 6\}$

$P_y(0) = \frac{1}{5}$ $P_y(2) = \frac{2}{5}$ $P_y(4) = \frac{1}{5}$ $P_y(6) = \frac{1}{5}$

→ $P_x(-1) + P_x(1) = \frac{2}{5}$

Law of Unconscious Statistician (LOTUS) for D.R.V.

$E[\underbrace{g(X)}_{Y}] = \sum_{x_k \in R_x} g(x_k) \cdot P_X(x_k)$

EX. prev. question → Find $E[Y]$ where $Y = 2|X|$

LOTUS $E[Y] = 2|-1| \cdot \frac{1}{5} + 2 \cdot |0| \cdot \frac{1}{5} + 2 \cdot |1| \cdot \frac{1}{5} + 2 \cdot |2| \cdot \frac{1}{5} + 2 \cdot |3| \cdot \frac{1}{5} = \frac{14}{5} = \mu_{E(y)}$

σ^2 Variance and σ standard deviation

- I am offered to invest $800. 2 investment accounts.

1- will give me $1000 in a year.

2- will give me either 500$ or $1500 (equally likely) in one year.

which one should I choose?

$P_x(1000) = 1$ → $EX = 1000$ $P_y(500) = \frac{1}{2}$ $P_y(1500) = \frac{1}{2}$

$EX = 500 \cdot \frac{1}{2} + 1500 \cdot \frac{1}{2} = 1000$ Same

Only expectation is not sufficient → sadece E yeterli değil

The variance of a RV X with $EX = \alpha_x$ $Var(x) = E[(x - \underbrace{\mu_x}_{mean})^2]$

→ $Var(x) = 1000 - 1000 = 0$

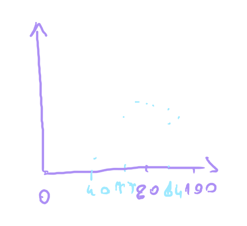
→ $Var(y) = \underbrace{500 - 1000}_{(-500)^2} + 1500 - 1000$?

"σ^2"

Standard deviation: is a square root of variance

$\sigma = std(X) = \sqrt{Var(x)}$

Ex:



100 students

mean → 78

variance → small

Covariance and Correlation

Ex, var, std. dev → distribution of a single R.V.

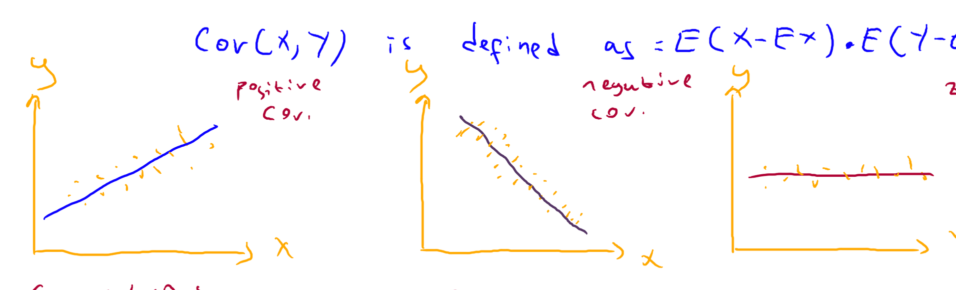
→ Two R.V. → covariance, correlation

→ Covariance measure the association of two R.V.

Covariance two R.V. X and Y

Cov(X,Y) is defined as $= E(X - EX) \cdot E(Y - EY) = E(XY) - E(X) \cdot E(Y)$

positive cov. negative cov. zero cov.



Cov(X,Y) > 0 both increasing

Cov(X,Y) < 0 → One of X,Y is inc. and the other one is decreasing

Cov(X,Y) = 0 → no relation btwn X,Y

Correlation

like covariance, the values of correlation is between [-1;1]

"ρ" $= \frac{Cov(x,y)}{std(x) \cdot std(y)} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$ rho

1,-1 Perfect correlation

$Cov(x,y) < 0$ small x, large Y

$Cov(x,y) = 0$ X and Y are uncorrelated

Properties of variances and covariances

a, b, c → constants

$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$
$Cov(aX + bY, cZ + dW)$
$= ac\,Cov(X,Z) + ad\,Cov(X,W) + bc\,Cov(Y,Z) + bd\,Cov(Y,W)$
$Cov(X,Y) = Cov(Y,X)$
$\rho(X,Y) = \rho(Y,X)$

In particular,
$Var(aX + b) = a^2 Var(X)$
$Cov(aX + b, cY + d) = ac\,Cov(X,Y)$
$\rho(aX + b, cY + d) = \rho(X,Y)$

For independent X and Y,
$Cov(X,Y) = 0$
$Var(X + Y) = Var(X) + Var(Y)$

Ex: A program consists two modules. The number of errors in the first module X and the number of errors in the second module Y have joint distribution

$P(x,y)$: $P(0,0) = P(0,1) = P(1,0) = 0.2$; $P(1,1) = P(1,2) = P(1,3) = 0.1$; $P(0,2) = P(0,3) = 0.05$

x \ y	0	1	2	3
0	0.2	0.2	0.05	0.05
1	0.2	0.1	0.1	0.1

a-) marginal Prob. X and Y P(x) P(y)

$P_x(0) = \sum_y P(0,y) = 0.2 + 0.2 + 0.05 + 0.05 = 0.5$ $P_x(1) = \sum_y P(1,y) = 0.2 + 0.1 + 0.1 + 0.1 = 0.5$

$P_y(0) = \sum_x P(x,0) = 0.2 + 0.2 = 0.4$ $P_y(1) = \sum_x P(x,1) = 0.2 + 0.1 = 0.3$ $P_y(2) = 0.05 + 0.1 = 0.15$

$P_y(3) = \sum_x P(x,3) = 0.15$

b-) X and Y are independent?

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

$P(0,0) = P_x(0) \cdot P_y(0)$

$0.2 \ne 0.5 \cdot 0.4$ → X and Y aren't independent

c-) Var(x)? Var(y)?

$Var(x) = E[(x - \mu_x)^2]$ $E(x) = \mu_x = \sum x P_x = 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5$

$Var(y) = E[(y - \mu_y)^2]$ $E(y) = \mu_y = \sum_{y=0}^{3} y P_y = 0 \cdot 0.4 + 1 \cdot 0.3 + 2 \cdot 0.15 + 3 \cdot 0.15 = 1.05$

$Var(x) = E[(x - \mu_x)^2] = \sum (x - 0.5)^2 \dots = 0.25$

$Var(y) = \sum_{y=0}^{3} (0 - 1.05)^2 \cdot 0.4 + (1 - 1.05)^2 \dots$

d-) $P(x,y) = ?$ covariance $\rho_{x,y} = cov(x,y) = E[(x - \mu_x)(y - \mu_y)]$

$= E(x \cdot y) - E(x) \cdot E(y)$

3.3.7 Chebyshev's inequality

Chebyshev's inequality: $P\{|X - \mu| > \varepsilon\} \le \left(\frac{\sigma}{\varepsilon}\right)^2$ for any distribution with expectation μ and variance σ^2 and for any positive ε.

Hence, $P\{|X - \mu| \ge \varepsilon\} \le \frac{\sigma^2}{\varepsilon^2}$.

Chebyshev's inequality shows that only a large variance may allow a variable X to differ significantly from its expectation μ. In this case, the risk of seeing an extremely low or extremely high value of X increases. For this reason, risk is often measured in terms of a variance or standard deviation.

Example 3.13. Suppose the number of errors in a new software has expectation $\mu = 20$ and a standard deviation of 2. According to (3.8), there are more than 30 errors with probability

$$P\{X > 30\} \le P\{|X - 20| > 10\} \le \left(\frac{2}{10}\right)^2 = 0.04.$$

However, if the standard deviation is 5 instead of 2, then the probability of more than 30 errors can only be bounded by $\left(\frac{5}{10}\right)^2 = 0.25$.

Chebyshev's inequality is universal because it works for any distribution. Often it gives a rather loose bound for the probability of $|X - \mu| > \varepsilon$. With more information about the distribution, this bound may be improved.

3.3.8 Application to finance

look chebyshevs inequality from book's book page 58