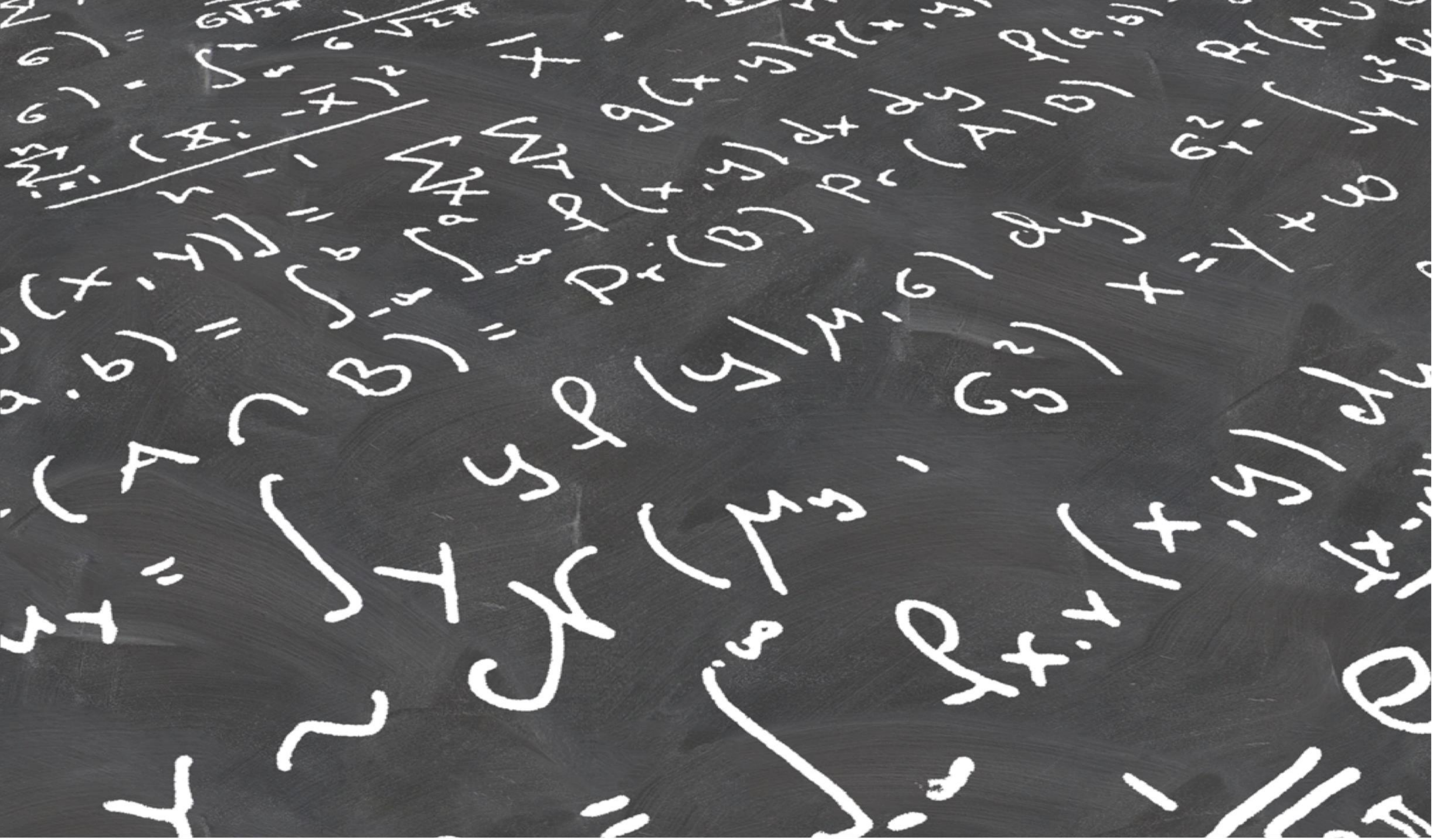


5. Simultaneous Random Variables and Transformations

Gunvor Elisabeth Kirkelund
Lars Mandrup

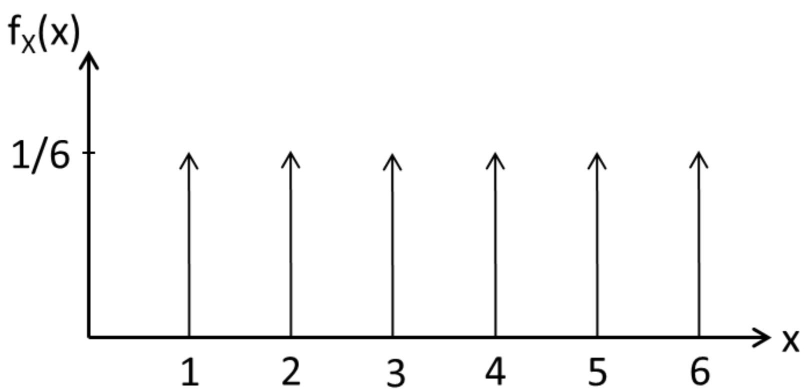
Agenda for Today

- Repetition:
  - One Random Variable
- Two Simultaneous Random Variables
- Joint pmf/pdf/cdf
- Correlation and Covariance
- Data sampling for test and simulation
- Transformation of random variables
- Sum of two random variables

One Stochastic Variable – Discrete

- Probability mass function (pmf):

$$f_X(x) = \begin{cases} Pr(X = x_i) & for\ X = x_i \\ 0 & otherwise \end{cases}$$

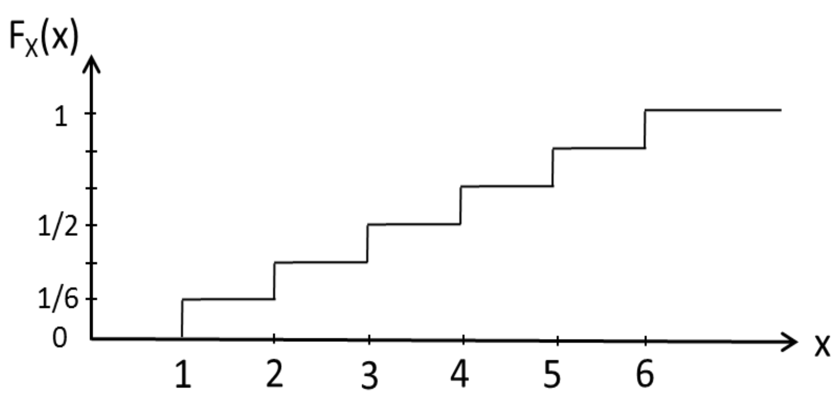


$$0 \leq f_X(x) \leq 1$$

$$\sum_{i=1}^{n} f_X(x_i) = \sum_{i=1}^{n} Pr(X = x_i) = 1$$

- Cumulative distribution function (cdf):

$$F_X(x) = Pr(X \leq x) = \sum_{i=1}^{n_x} f_X(x_i)$$



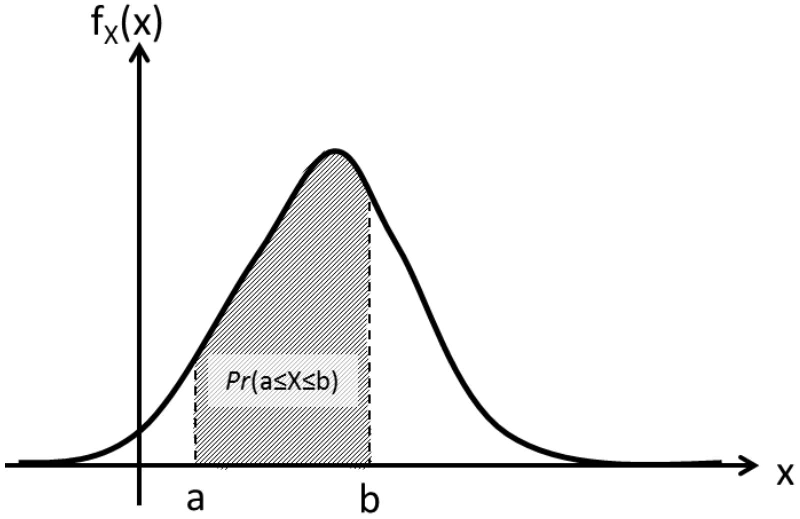
$$0 \leq F_X(x) \leq 1$$

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

One Stochastic Variable – Continuous

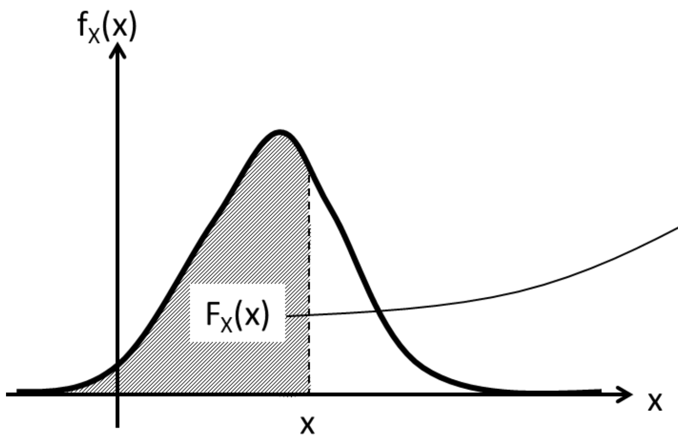
- Probability density function (pdf): $$Pr(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

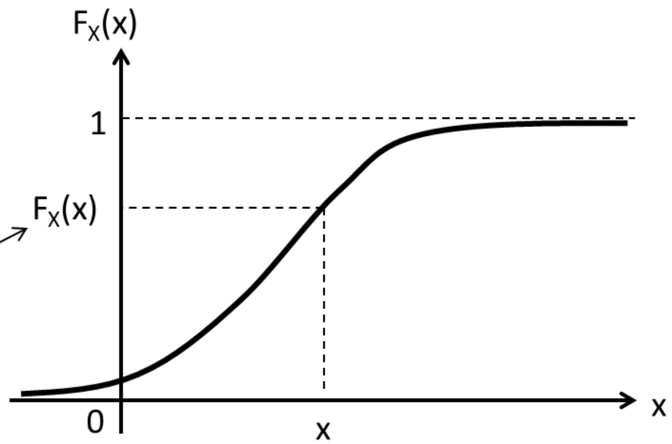


$$f_X(x) \geq 0$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

- Cumulative distribution function (cdf): $$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = Pr(X \leq x)$$





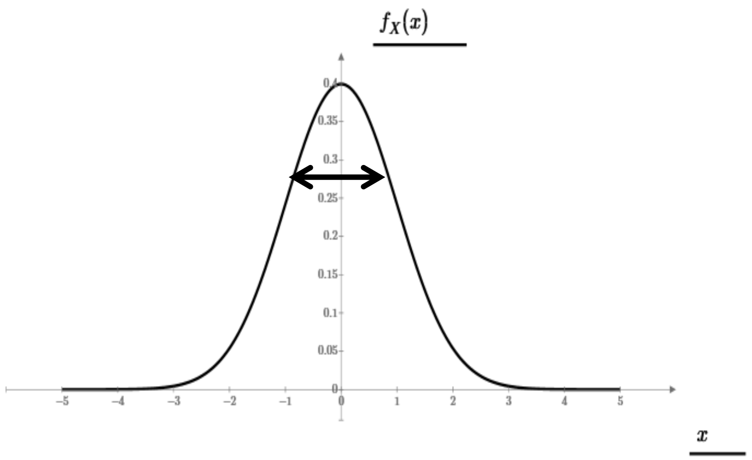
$$0 \leq F_X(x) \leq 1$$

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

Expectation, Variance and Standard Deviation

- Mean value: $EX = E[X] = \bar{X} = \mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x)dx \quad (\sum_{i=1}^{n} x_i \, f_X(x_i))$
- Variance: $Var(\mathrm{X}) = \sigma_X{}^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f_X(x)dx = E\left[X^2\right] - E[X]^2$

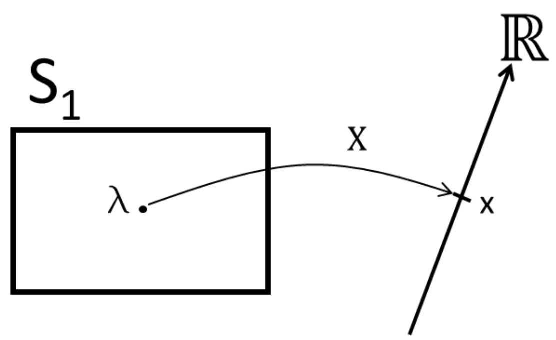


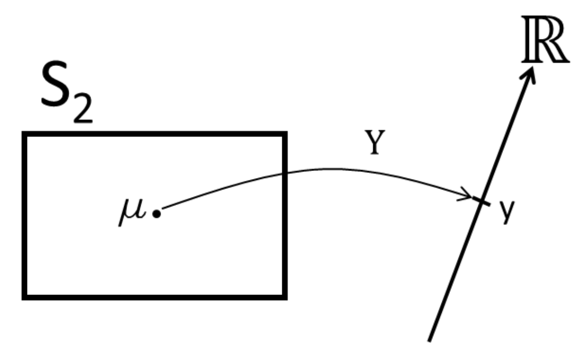
- Standard deviation: $\sigma_X = \sqrt{Var(\mathrm{X})}$
- A function: $E[g(X)] = \overline{g(X)} = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx \quad (\sum_{i=1}^{n} g(x_i) \, f_X(x_i))$

  $Var\big(g(\mathrm{X})\big) = \int_{-\infty}^{\infty} \left(g(x) - \overline{g(x)}\right)^2 \cdot f_X(x)dx = E\left[g(X)^2\right] - E[g(X)]^2$
- Linear function: $E[aX + b\,] = a \cdot E[X] + b$

  $Var[aX + b\,] = a^2\left(E[X^2] - E[X]^2\right) = a^2 \cdot Var(X)$

Two Simultaneous Discrete Random Variables



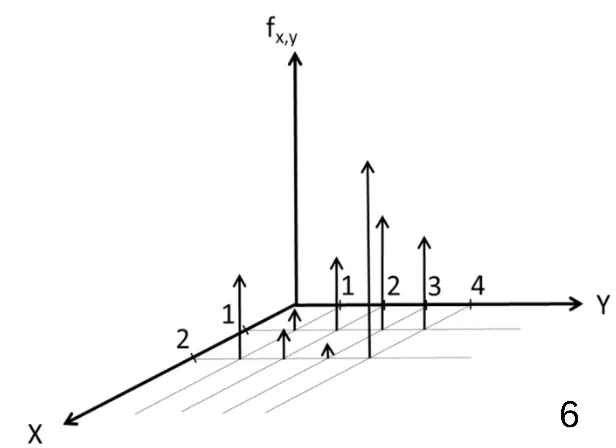


- Two (or more) discrete random variables X and Y
- We can discribe the two probabilities as a simultaneous pmf:

Joint (Simultaneous) pmfs:

$$f_{X,Y}(x,y) = \begin{cases} Pr\Big((X = x_i\ ) \cap (Y = y_j)\Big) & for\ \ X = x_i\ \wedge\ Y = y_j \\ 0 & otherwise \end{cases}$$

➢ $0 \leq f_{X,Y}(x,y) \leq 1$ ➢ $\sum_x \sum_y f_{X,Y}(x,y) = 1$

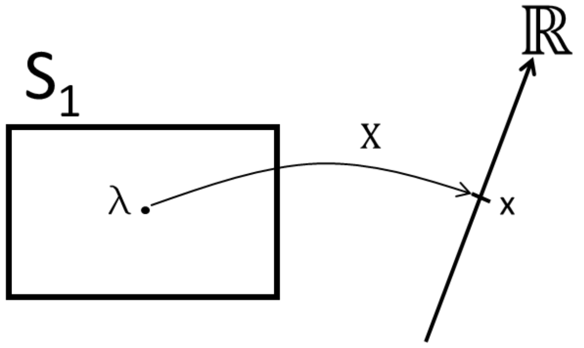


Fx.: X = The number of bicycles in front of IHA
Y = The number of people inside IHA

Two Simultaneous Discrete Random Variables

Cumulative Distribution Function cdf:

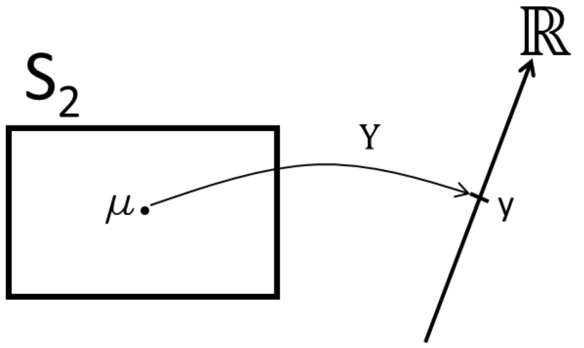
$$F_{X,Y}(x,y) = Pr(X \le x \wedge Y \le y) = \sum_{x_i \le x} \sum_{y_i \le y} f_{X,Y}(x_i, y_i)$$



- $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$
- $F_{X,Y}(\infty, \infty) = 1$

Marginal pmfs:

$$f_X(x) = \sum_{y} f_{X,Y}(x,y) \qquad f_Y(y) = \sum_{x} f_{X,Y}(x,y)$$

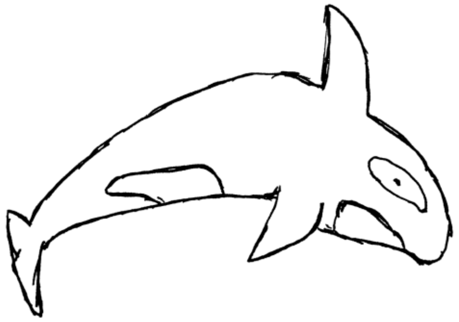


Conditional pmfs / Bayes Rule:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = Pr(X = x|\mathrm{Y} = \mathrm{y})$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = Pr(Y = y|\mathrm{X} = \mathrm{x})$$

Orca Example



- In a conversation effort, we look for dead orcas when we are visiting an ocean.
- Stochastic variables:
  - X = Gender: $R_X = \{1,2\}$
  - Y = Location: $R_Y = \{1,2,3,4\}$

Gender (X) \ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)
Female (1)	2	7	11	9
Male (2)	8	3	1	19

- Total number of observed dead orcas = 60.

Orca Example – joint and marginal pmf

- The discrete simultaneous mass function (pmf) for observing a orca at a specific ocean and its gender is

$f_{X,Y}(x,y)$ $f_X(x)$

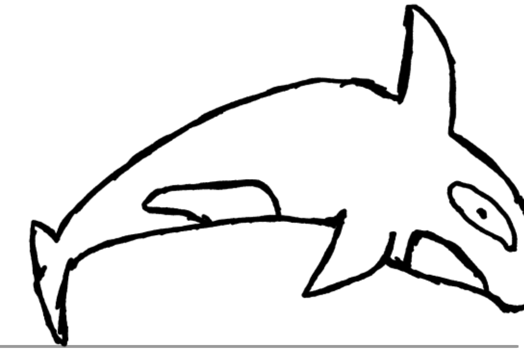
Gender (X) \ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total
Female (1)	2/60	7/60	11/60	9/60	29/60
Male (2)	8/60	3/60	1/60	19/60	31/60
Total	10/60	10/60	12/60	28/60	1

$f_Y(y)$

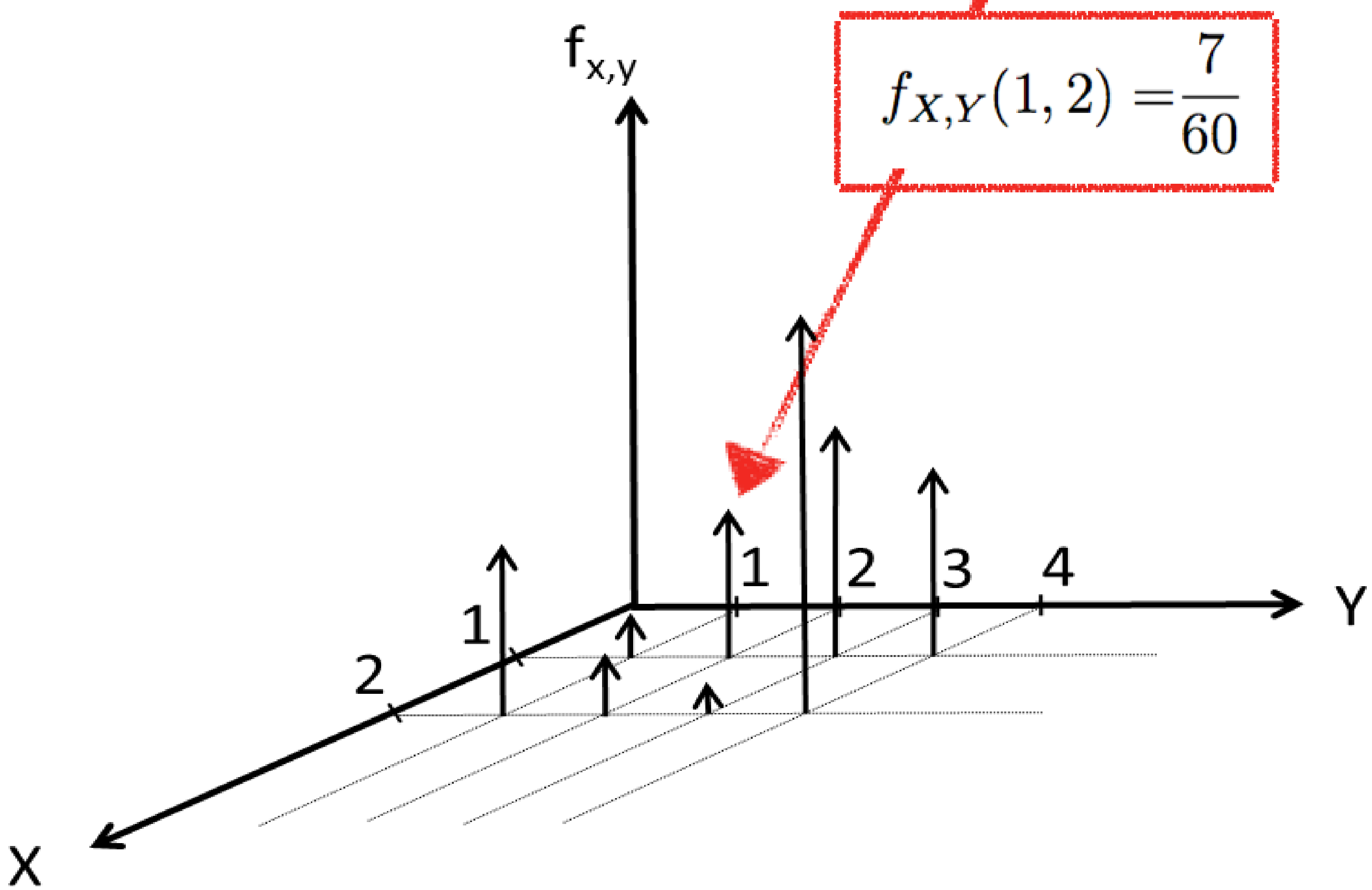
$$f_X(1) = f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(1,3) + f_{X,Y}(1,4) = \frac{2}{60} + \frac{7}{60} + \frac{11}{60} + \frac{9}{60} = \frac{29}{60}$$

$$f_X(2) = f_{X,Y}(2,1) + f_{X,Y}(2,2) + f_{X,Y}(2,3) + f_{X,Y}(2,4) = \frac{8}{60} + \frac{3}{60} + \frac{1}{60} + \frac{19}{60} = \frac{31}{60}$$

Orca Example - Joint pmf



Gender (X) \ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total f_X
Female (1)	2/60	7/60	11/60	9/60	29/60
Male (2)	8/60	3/60	1/60	19/60	31/60
Total f_Y	10/60	10/60	12/60	28/60	1



$$f_{X,Y}(1,2) = \frac{7}{60}$$

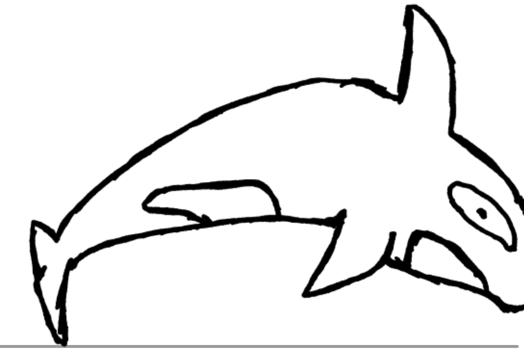
Fx.:

$$Pr(Female|Antartica)$$

$$= f_{X|Y}(1|2) = \frac{f_{X,Y}(1,2)}{f_Y(2)}$$

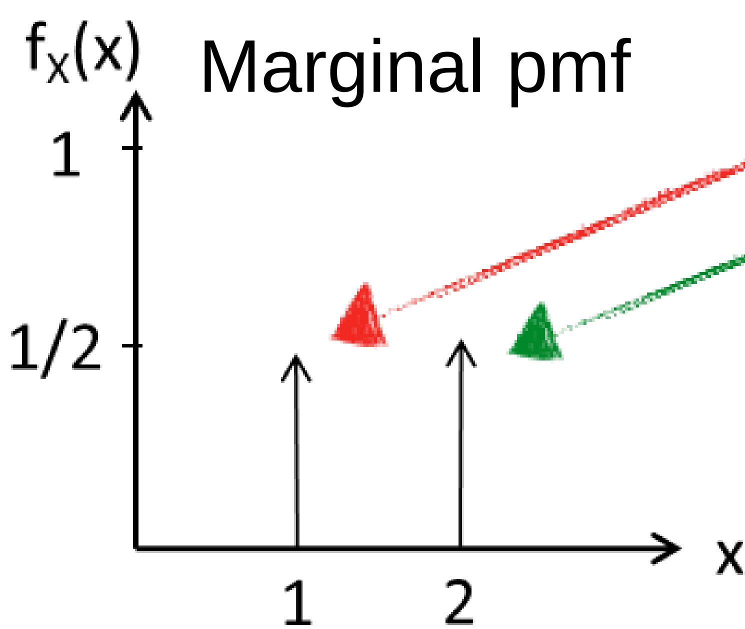
$$= \frac{7/60}{10/60} = \frac{7}{10} = 0{,}7$$

Orca Example – Marginal pmf and cdf



Gender (X) \ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total f_X
Female (1)	2/60	7/60	11/60	9/60	$\boxed{29/60}$
Male (2)	8/60	3/60	1/60	19/60	$\boxed{31/60}$
Total f_Y	10/60	10/60	12/60	28/60	1

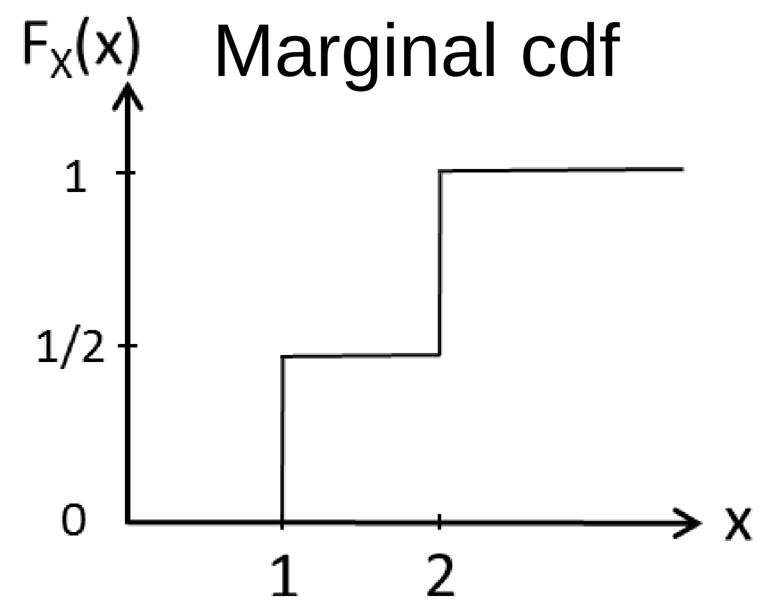
Marginal pmf



$$f_X(1) = \frac{29}{60}$$

$$f_X(2) = \frac{31}{60}$$

Marginal cdf

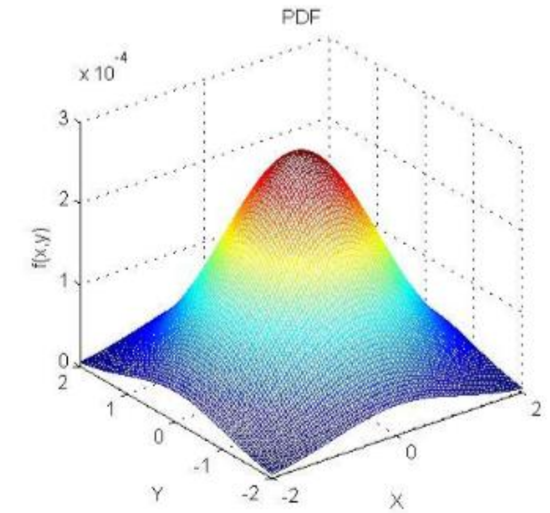


$$F_X(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{29}{60} & \text{for } 1 \le x < 2 \\ 1 & \text{for } 2 \le x \end{cases}$$

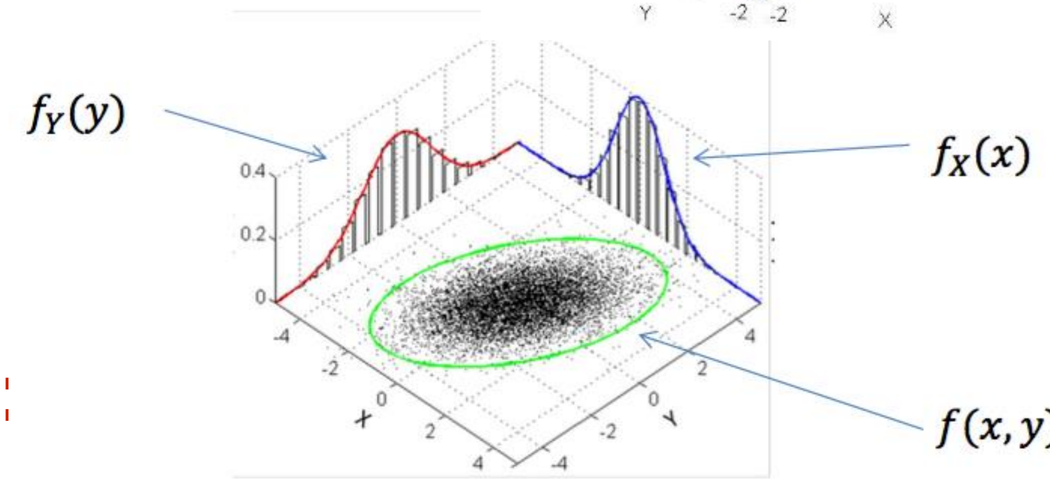
Two Simultaneous Continuous Random Variables

Joint (Simultaneous) pdf: $f_{X,Y}(x,y) \geq 0$ $\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

$$Pr\big((a \leq X \leq b) \cap (c \leq Y \leq d)\big) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy$$



Marginals:
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$$



Cumulative Distribution Function cdf:

cdf $\quad F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(x,y)dxdy = Pr(X \leq x \wedge Y \leq y)$

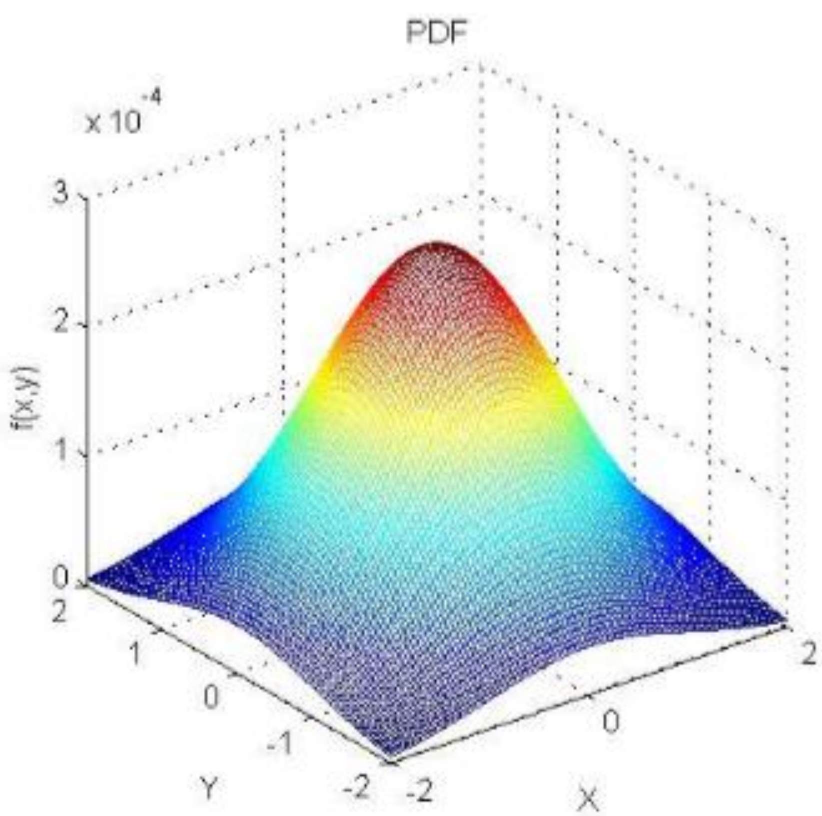
pdf $\quad f_{X,Y}(x,y) = \dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$

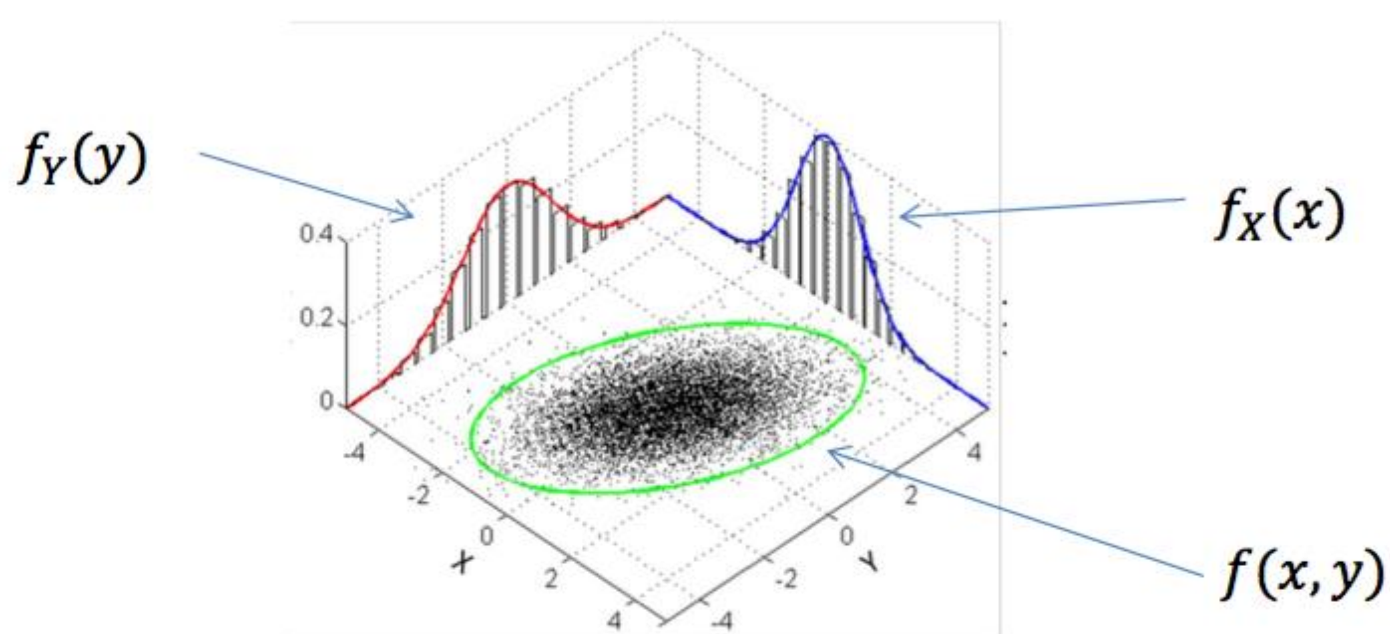
Bivariate (2D) Normal Distribution

Two dimensional Gaussian $$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{z}{2(1-\rho^2)}\right)$$

$$z = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}$$

$$\rho = \frac{E[XY]-E[X]E[Y]}{\sigma_X\sigma_Y}$$ Correlation coefficient





Independence

- We have independence between X and Y if and only if:

$$\boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)} \quad \text{for all } x \text{ and } y$$

Example of independent random variables:

- A persons height and the current exact distance from the earth to the moon.

Example of dependent random variables:

- The time of day and the amount of bicycles parked the at the engineering college.
- The energy of a mobile signal and the length in meters to a basestation.

Bayes Rule and Independence

Independence: $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ for all x and y

- Bayes Rule: $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

  gives that if X and Y are independent, then:

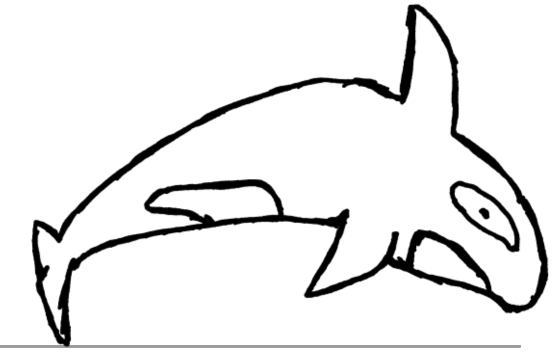
$$\boxed{f_{X|Y}(x|y) = f_X(x)}$$

- Also:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \Rightarrow E[XY] = E[X]E[Y]$$

  - but the opposite is not allways true!

Orca Example – Independence



Gender (X) \ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total f_X
Female (1)	2/60	7/60	11/60	9/60	29/60
Male (2)	8/60	3/60	1/60	19/60	31/60
Total f_Y	10/60	10/60	12/60	28/60	1

Fx: $f_X(1) = \frac{29}{60}, \quad f_Y(3) = \frac{12}{60}$

$$f_{XY}(1,3) = \frac{11}{60} \neq \frac{29}{60} \cdot \frac{12}{60} = f_X(1) \cdot f_Y(3)$$

→ Gender (X) and Location (Y) are not independent!

Expectations of Simultaneous Random Variables

Expectation (mean) of Simultaneous Random Variables:

- $\boxed{E[X|Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = h(Y)}$ $\quad (\sum_{x_i} x_i f_{X|Y}(x_i|y_i))$

- $\boxed{E[g(X)|Y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx = h_g(Y)}$ ← LOTUS

Law of Total (Iterated) Expectation:

- $$\boxed{\begin{aligned} EX = E\big[E[X|Y]\big] &= \int_{-\infty}^{\infty} E[X|Y] \cdot f_Y(y)dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y) \cdot f_Y(y)dydx \end{aligned}}$$

$(\sum_{y_i}\sum_{x_i} x_i f_{X|Y}(x_i|y_i) \cdot f_Y(y_i))$

Orca Example – Expectation

Gender (X) \ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total f_X
Female (1)	2/60	7/60	11/60	9/60	29/60
Male (2)	8/60	3/60	1/60	19/60	31/60
Total f_Y	10/60	10/60	12/60	28/60	1

$$E[Y|X] = \sum_{y=1}^{4} y \cdot f_{Y|X}(y|x) = \sum_{y=1}^{4} y \cdot \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \begin{cases} \sum_{y=1}^{4} y \cdot \frac{f_{X,Y}(1,y)}{f_X(1)} = 1 \cdot \frac{f_{X,Y}(1,1)}{f_X(1)} + 2 \cdot \frac{f_{X,Y}(1,2)}{f_X(1)} + 3 \cdot \frac{f_{X,Y}(1,3)}{f_X(1)} + 4 \cdot \frac{f_{X,Y}(1,4)}{f_X(1)} & for\ x = 1 \\ \sum_{y=1}^{4} y \cdot \frac{f_{X,Y}(2,y)}{f_X(2)} = 1 \cdot \frac{f_{X,Y}(2,1)}{f_X(2)} + 2 \cdot \frac{f_{X,Y}(2,2)}{f_X(2)} + 3 \cdot \frac{f_{X,Y}(2,3)}{f_X(2)} + 3 \cdot \frac{f_{X,Y}(2,4)}{f_X(2)} & for\ x = 2 \end{cases}$$

$$= \begin{cases} 1 \cdot \frac{2/60}{29/60} + 2 \cdot \frac{7/60}{29/60} + 3 \cdot \frac{11/60}{29/60} + 4 \cdot \frac{9/60}{29/60} = \frac{85}{29} = 2{,}93 & for\ x = 1 \\ 1 \cdot \frac{8/60}{31/60} + 2 \cdot \frac{3/60}{31/60} + 3 \cdot \frac{1/60}{31/60} + 4 \cdot \frac{19/60}{31/60} = \frac{93}{31} = 3{,}00 & for\ x = 2 \end{cases} = h(x)$$

Orca Example – Total Expectation

Gender (X) \ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total f_X
Female (1)	2/60	7/60	11/60	9/60	29/60
Male (2)	8/60	3/60	1/60	19/60	31/60
Total f_Y	10/60	10/60	12/60	28/60	1

$$EY = E\big[E[Y|X]\big] = \sum_{x=1}^{2} E[Y|X] \cdot f_X(x) = 2{,}93 \cdot \frac{29}{60} + 3{,}00 \cdot \frac{31}{60} = \frac{91}{60} = 2{,}97$$

$$EY = \sum_{y=1}^{4} y \cdot f_Y(y) = 1 \cdot \frac{10}{60} + 2 \cdot \frac{10}{60} + 3 \cdot \frac{12}{60} + 4 \cdot \frac{28}{60} = \frac{178}{60} = 2{,}97$$

$$EX = E\big[E[X|Y]\big] = \sum_{y=1}^{4} E[X|Y] \cdot f_Y(y) = \sum_{x=1}^{2} x \cdot f_X(x) = 1 \cdot \frac{29}{60} + 2 \cdot \frac{31}{60} = \frac{91}{60} = 1{,}52$$

Tells of the coupling between variables

Correlation and Covariance

Correlation tells of the (biased) coupling between variables

- Correlation: $corr(X,Y) = E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x,y)dxdy$

Covariance is without bias from the mean

- Covariance: $cov(X,Y) = E\left[(X-\bar{X})(Y-\bar{Y})\right] = E[XY] - E[X] \cdot E[Y]$

Correlation Coefficient is the normalized Covariance

- Correlation coefficient: $\rho = E\left[\frac{X-\bar{X}}{\sigma_X} \cdot \frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY] - E[X]E[Y]}{\sigma_X \cdot \sigma_Y}$

$$-1 \leq \rho \leq 1$$

Correlation tells of the coupling between variables

Correlation Coefficient

- The correlation coefficient, is an indicator on how much two random variables X and Y are correlated.

$$\boxed{\rho = E\left[\frac{X-\bar{X}}{\sigma_X}\cdot\frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY]-E[X]E[Y]}{\sigma_X\sigma_Y}}$$

- We have that: $-1 \leq \rho \leq 1$

Uncorrelated

- If X and Y are independent:

$$E[XY] = E[X]\cdot E[Y] \quad \text{and} \quad cov(X,Y) = \rho = 0$$

  - ➢ but the opposite is not allways true!

- If $\rho = 1$: $Y = aX + b; \quad a > 0$ (Max) positively correlated
- If $\rho = -1$: $Y = aX + b; \quad a < 0$ (Max) negatively correlated

Orca Example – Correlation

Gender (X) \ Location (Y)	**Atlantic (1)**	**Antartica (2)**	**Pacific (3)**	**Seaworld (4)**	**Total** f_X
Female (1)	2/60	7/60	11/60	9/60	29/60
Male (2)	8/60	3/60	1/60	19/60	31/60
Total f_Y	10/60	10/60	12/60	28/60	1

$EX = \frac{91}{60} = 1{,}52;\quad E\left[X^2\right] = \sum_{x=1}^{2} x^2 \cdot f_X(x) = \frac{153}{60} = 2{,}55;\quad \sigma_X^2 = E\left[X^2\right] - EX^2 = \frac{899}{3600} = 0{,}25;$

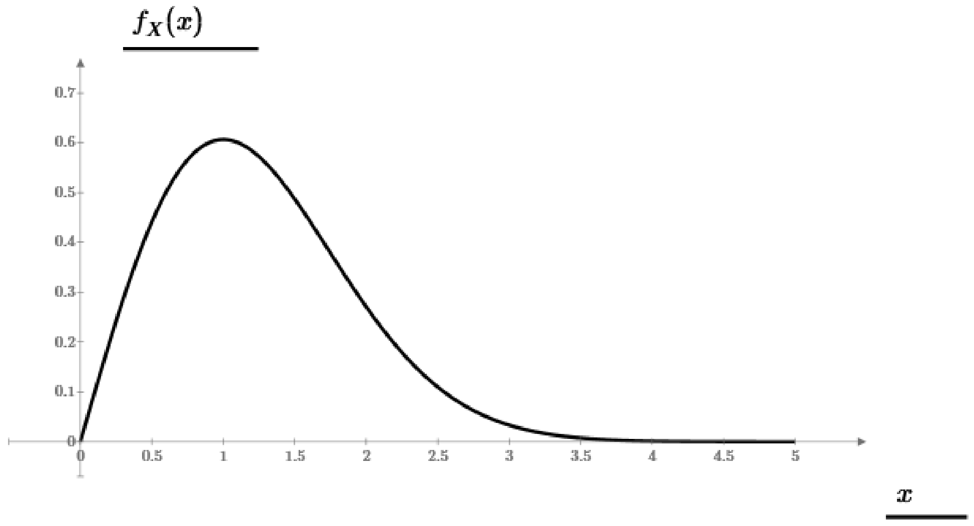
$EY = \frac{178}{60} = 2{,}97;\quad E\left[Y^2\right] = \sum_{y=1}^{4} y^2 \cdot f_Y(y) = \frac{606}{60} = 10{,}1;\quad \sigma_Y^2 = E\left[Y^2\right] - EY^2 = \frac{4676}{3600} = 1{,}30;$

$$E[XY] = \sum_{x=1}^{2}\sum_{y=1}^{4} x \cdot y \cdot f_{X,Y}(x,y) = 1 \cdot 1 \cdot \frac{2}{60} + 1 \cdot 2 \cdot \frac{7}{60} + \cdots + 2 \cdot 4 \cdot \frac{19}{60} = \frac{271}{60} = 4{,}52$$

- $corr(X,Y) = E[XY] = 4{,}52$
- $cov(X,Y) = E[XY] - EX \cdot EY = \frac{271}{60} - \frac{91}{60} \cdot \frac{178}{60} = \frac{62}{3600} = 0{,}0172$
- $\rho = \frac{E[XY] - E[X]E[Y]}{\sigma_X . \sigma_Y} = \frac{0{,}0172}{\sqrt{0{,}25} \cdot \sqrt{1{,}30}} = 0{,}030 \neq 0$ → X and Y are not independent

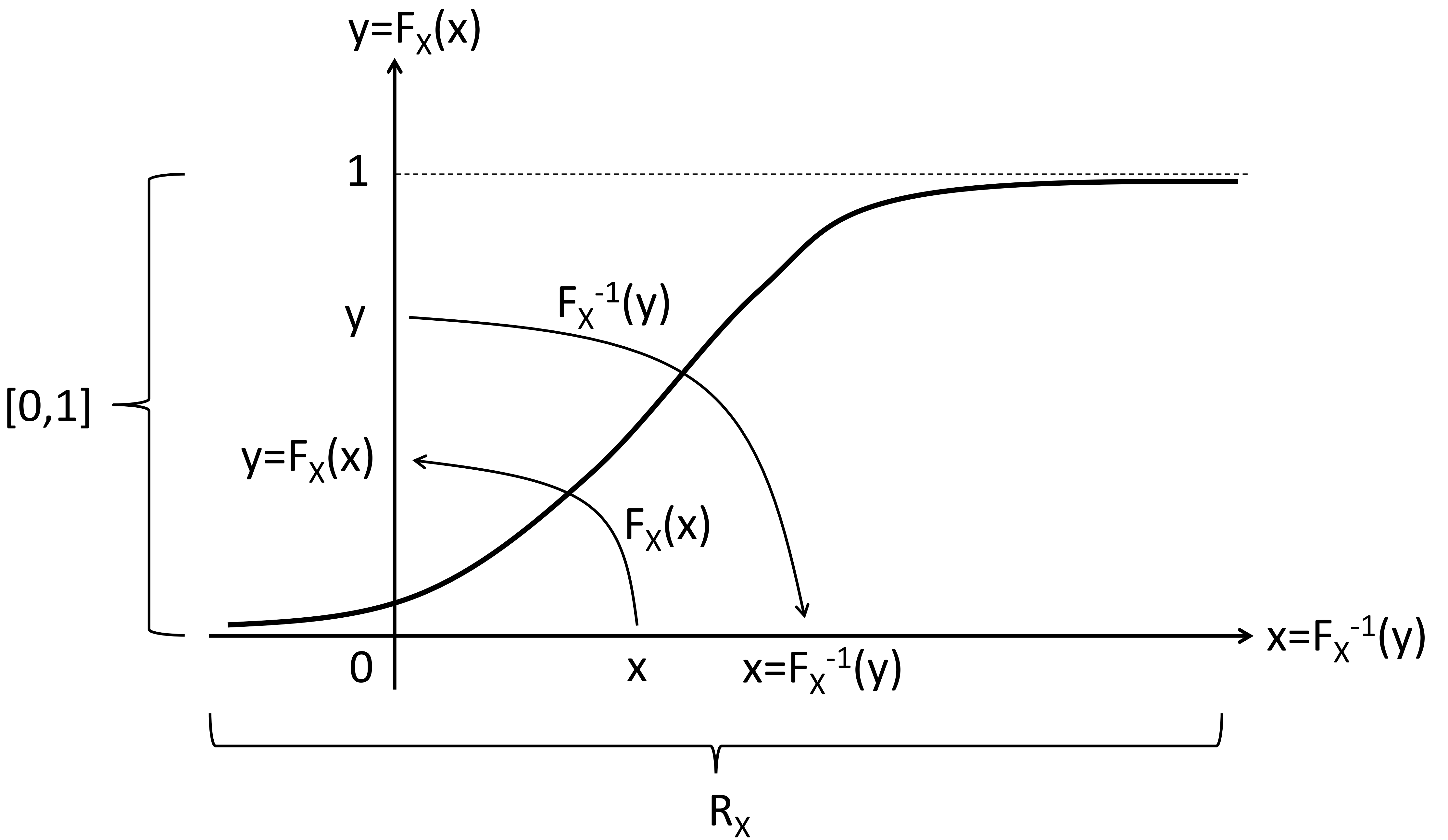
Sampling Random Testdata

- In a flight simulator, the altitude of the plane is simulated to be Rayleigh distributed.
- For a given initial height, draw a Rayleigh distributed sample.



Sampling From Any Distribution

For test or simulation you need testdata (”measurements”) randomly sampled from a given distribution:



Sampling From Any Distribution

For test or simulation you need testdata (”measurements”) randomly sampled from a given distribution:

- Standard distributions: $X \sim \mathcal{N}\left(\mu, \sigma^2\right) \to randn()$
  $X \sim \mathcal{U}(a, b) \to rand()$

- Other distributions:
  - Find the cdf of the distribution: $F_X(x)$
  - Find the inverse of the cdf: $y = F_X(x) \;\Rightarrow\; x = F_X^{-1}(y)$
  - Draw a random sample: $y \sim \mathcal{U}[0; 1]$
  - Insert into the inverse cdf: $x = F_X^{-1}(y)$
  - The samples $X = x$ is distributed according to: $F_X(x)$

Transformation of Stochastic Variable X to Y

- Given:
  - pdf/pmf: $f_X(x)$
  - Function/Transformation: $\mathrm{Y} = g(X)$ (Fx: $Y = X^2$)

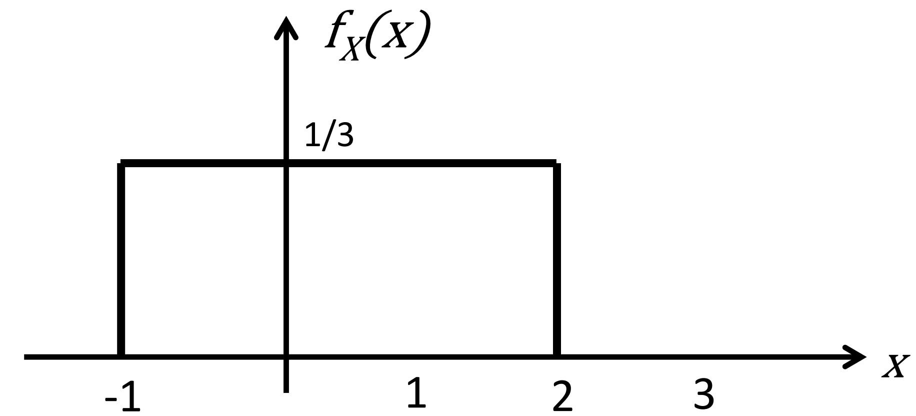
- Find new pmf: $f_Y(y)$:

➢ Discrete: $f_Y(y) = Pr(Y = y) = Pr(g(X) = y) = \sum_{x:g(x)=y} f_X(x)$

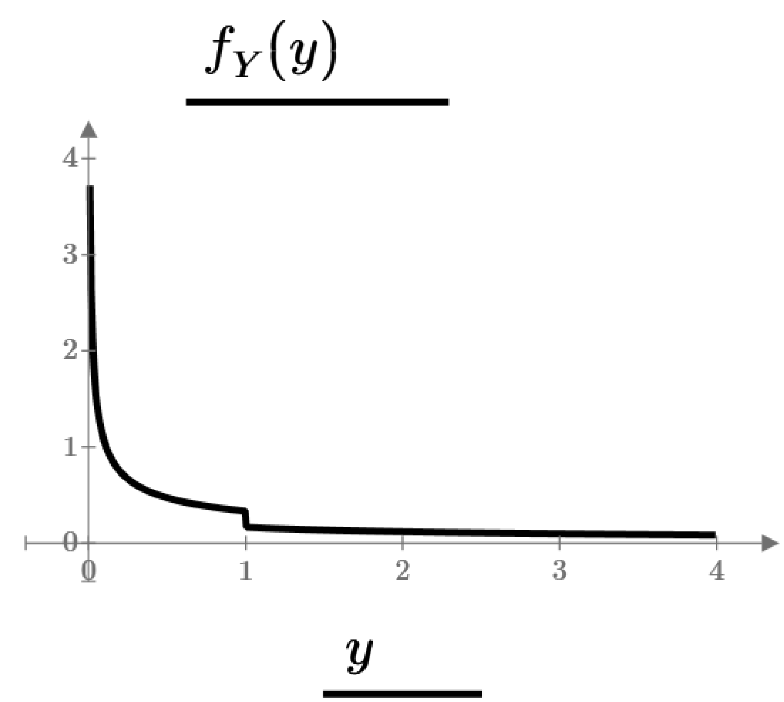
➢ Continuous: $f_Y(y) = \sum \left|\frac{dx(y)}{dy}\right| f_X\big(g^{-1}(y)\ \big) = \sum_{x:g(x)=y} \left(\frac{f_X(x)}{|g\prime(x)|}\right)$

Example: Transformation of Continuous Random Variable

- $X \sim \mathcal{U}(-1;2)$
- $Y = X^2 = g(X)$



- $f_Y(y) = \begin{cases} \frac{1}{3\cdot\sqrt{y}}; & 0 < y < 1 \\ \frac{1}{6\cdot\sqrt{y}}; & 1 < y < 4 \end{cases}$



Distribution of the Sum of Two Random Variables

- Let: $Z = X + Y$

  where X and Y are two random variables X and Y with density functions $f_X(x)$ and $f_Y(y)$

- Then: $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx = \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)dy$

- If X and Y are independent:

- $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(z-y) \cdot f_Y(y)dy$

  $= f_X(z) * f_Y(z)$

Convolution of Two functions

Mean and Variance of the Sum of Two Random Variables

For a random variable: $Z = X + Y$

- $\boxed{EZ = EX + EY}$

X and Y can both be dependent and independent

- $\boxed{Var(Z) = Var(X) + Var(Y) + 2cov(X, Y)}$

  where: $cov(X, Y) = E[XY] - E[X]E[Y]$

- OBS: If X and Y are independent: $cov(X, Y) = 0$ and so:

$$\boxed{Var(Z) = Var(X) + Var(Y)}$$

Two Random Variables

Two random variables: X and Y

- Simultaneous pdf: $f_{X,Y}(x,y)$
- Marginal pdf: $f_X(x)$ and $f_Y(y)$
- Conditional pdf: $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$
- Simultaneous cdf: $F_{X,Y}(x,y)$
- Correlation: $corr(X,Y) = E[XY]$
- Covariance: $cov(X,Y) = E[XY] - E[X]E[Y]$
- Correlation coefficient: $\rho = \frac{E[XY]-E[X]E[Y]}{\sigma_X \cdot \sigma_Y}$

- Sum: $Z = X + Y$
- Expectation: $EZ = EX + EY$
- Variance: $\mathrm{Var}(Z) = Var(X) + Var(Y)$ if <u>independent</u>
  $\mathrm{Var}(Z) = Var(X) + Var(Y) + 2cov(X,Y)$ if <u>dependent</u>

Words and Concepts to Know

Marginal

Simultaneous Random Variables

Convolution

Joint pmf

Correlation

Transformation of stochastic variables

Rayleigh Distribution

Correlation coefficient

Joint pdf

Simultaneous pmf

Randomly Sampled Data

Covariance

Simultaneous pdf

Bivariate Normal Distribution