Review: mostly probability and some statistics

C2

Content

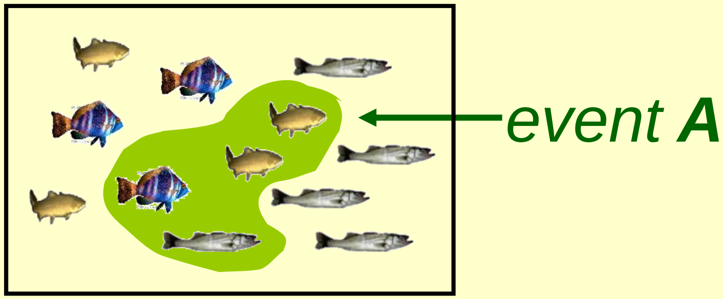
- Probability (should know already)
  - Axioms and properties
  - Conditional probability and independence
  - Law of Total probability and Bayes theorem
- Random Variables
  - Discrete
  - Continuous
- Pairs of Random Variables
- Random Vectors
- Gaussian Random Variable

Basics

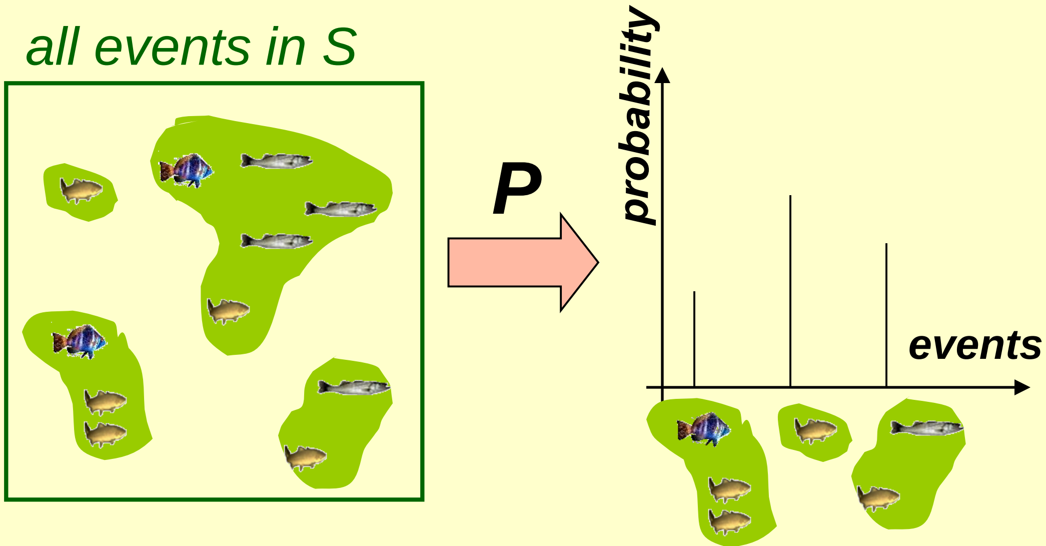
- We are performing a random experiment (catching fish from the sea)
- Sample space *S*: the set of all possible outcomes
- An event *A:* a set of possible outcomes of experiment, i.e. a subset of *S*
- *Probability law:a* rule *that* assigns probabilities to events in an experiment

$$A \longrightarrow \boldsymbol{P}(A)$$

S: *all fish in the sea*



total number of events: 2^{12}



Axioms of Probability

1. $P(A) \geq 0$
2. $P(S) = 1$
3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

Properties of Probability

$$P(\varnothing) = 0$$

$$P(A) \leq 1$$

$$P(A^c) = 1 - P(A)$$

$$A \subset B \ \Rightarrow P(A) < P(B)$$

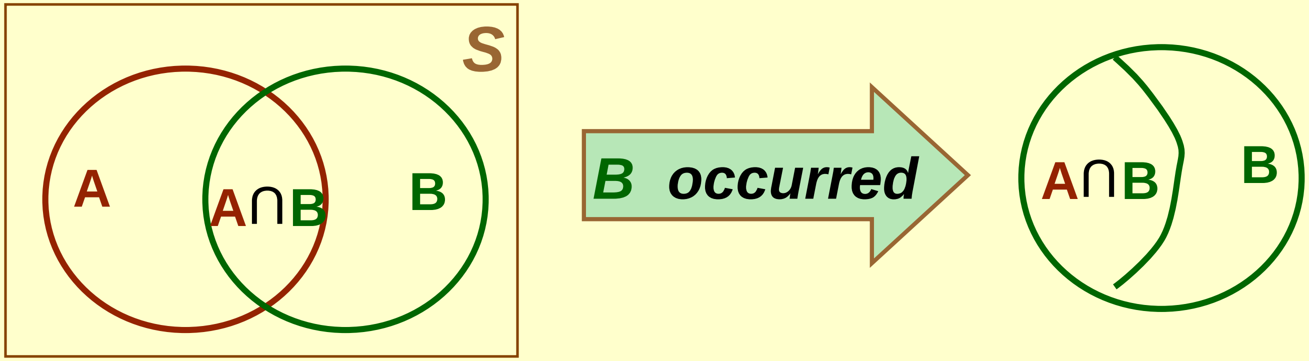
$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$$

$$\left\{A_i \cap A_j = \varnothing, \forall i, j\right\} \Rightarrow P\left(\bigcup_{k=1}^{N} A_k\right) = \sum_{k=1}^{N} P\left(A_k\right)$$

Conditional Probability

- If A and B are two events, and we know that event B has occurred, then (if P(B)>0)

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$



the "new" sample space is **B**, the "new" **A** is old **A∩B**

- multiplication rule $P(A\cap B)=P(A|B)\,P(B)$

Independence

- A and B are independent events if

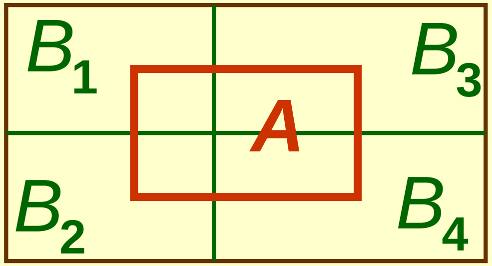
$$P(A \cap B) = P(A)\,P(B)$$

- By the law of conditional probability, if A and B are independent

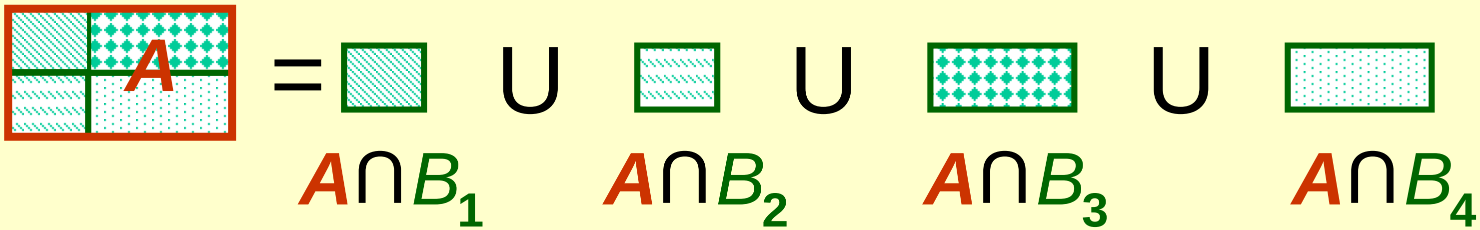
$$P(A|B) = \frac{P(A)\,P(B)}{P(B)} = P(A)$$

- If two events are not independent, then they are said to be dependent

Law of Total Probability



- $B_1, B_2, \ldots, B_n$ partition S
- Consider an event A



- Thus $P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + P(A \cap B_4)$
- Or using multiplication rule:

$$P(A) = P(A \mid B_1)P(B_1) + \ldots + P(A \mid B_4)P(B_4)$$

$$P(A) = \sum_{k=1}^{n} P(A \mid B_k)P(B_k)$$

Bayes Theorem

- Let B_1, B_2, …, B_n, be a partition of the sample space S. Suppose event A occurs. What is the probability of event B_i ?
- **Answer: Bayes Rule**

from conditional probability

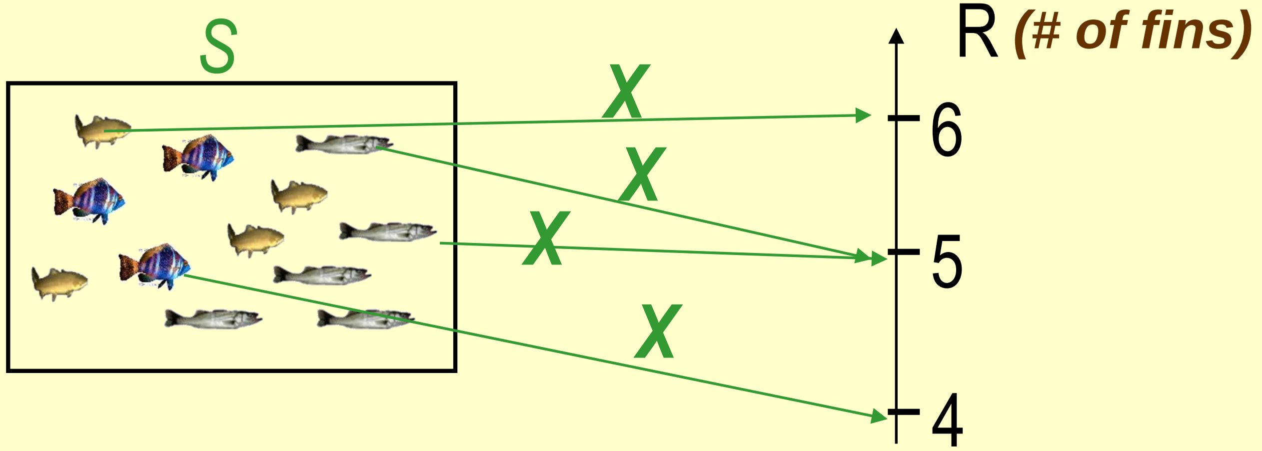
$$P(B_i \mid A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A \mid B_i)P(B_i)}{\sum_{k=1}^{n} P(A \mid B_k)P(B_k)}$$

from the law of total probability

- One of the most useful tools we are going to use

Random Variables

- A random variable ***X*** is a function from sample space ***S*** to a real number. ***X***: ***S*** ⟶ R



- ***X*** is random due to randomness of its argument
- $$P(X = a) = P(X(\omega) = a) = P(\omega \mid X(\omega) = a)$$

Two Types of Random Variables

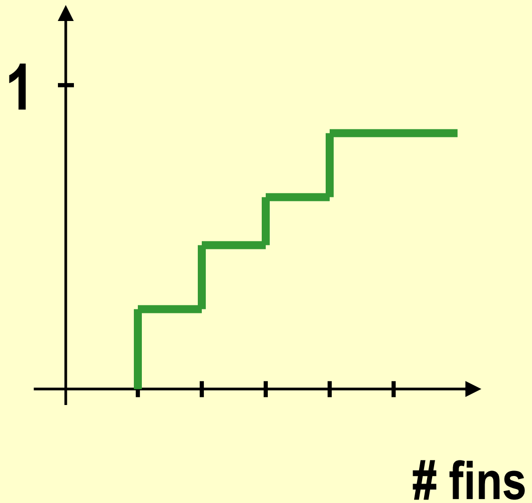
- ***Discrete*** random variable has countable number of values
  - number of fish fins (0,1,2,….,30)
- ***Continuous*** random variable has continuous number of values
  - fish weight (any real number between 0 and 100)

Cumulative Distribution Function

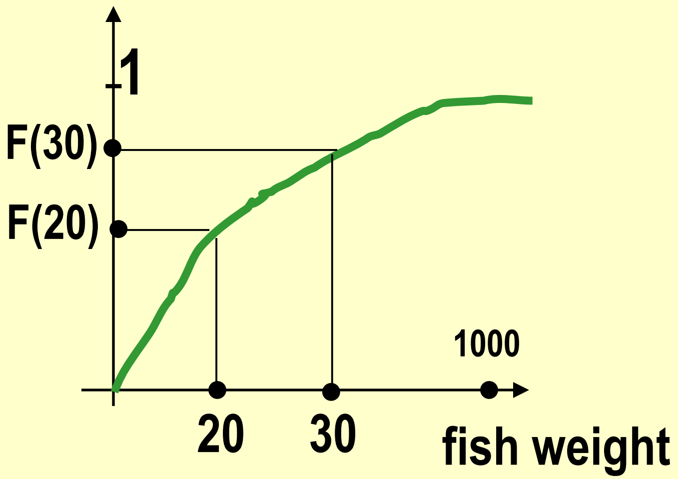
- Given a random variable ***X***, CDF is defined as

$$F(a) = P(X \leq a)$$

CDF for discrete rv

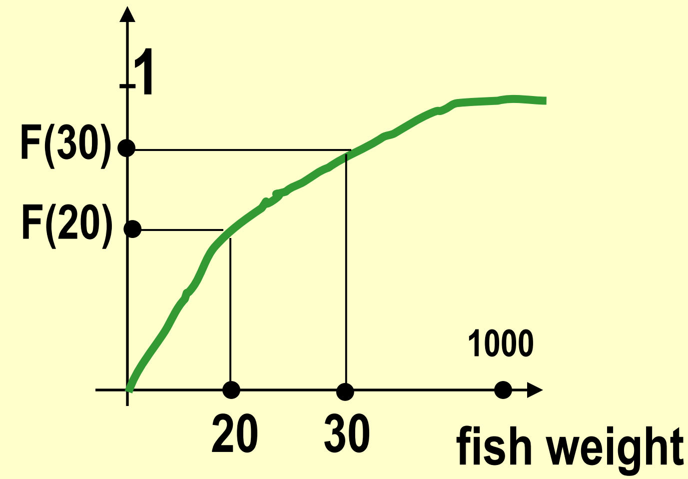


CDF for continuous rv



Properties of CDF $F(a) = P(X \le a)$

CDF for continuous rv



1. $F(a)$ is non decreasing
2. $\lim_{b \to \infty} F(b) = 1$
3. $\lim_{b \to -\infty} F(b) = 0$

- Questions about ***X*** can be asked in terms of CDF

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

Example:

P(fish weights between 20 and 30)=F(30)-F(20)

Discrete RV: Probability Mass Function

- Given a discrete random variable ***X***, we define the probability mass function as

$$p(a) = P(X = a)$$

- Satisfies all axioms of probability

- CDF in discrete case satisfies

$$F(a) = P(X \le a) = \sum_{x \le a} P(X = a) = \sum_{x \le a} p(a)$$

Continuous RV: Probability Density Function

- Given a continuous RV ***X***, we say f(x) is its probability density function if
  - $F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$
  - and, more generally $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx$

Properties of Probability Density Function

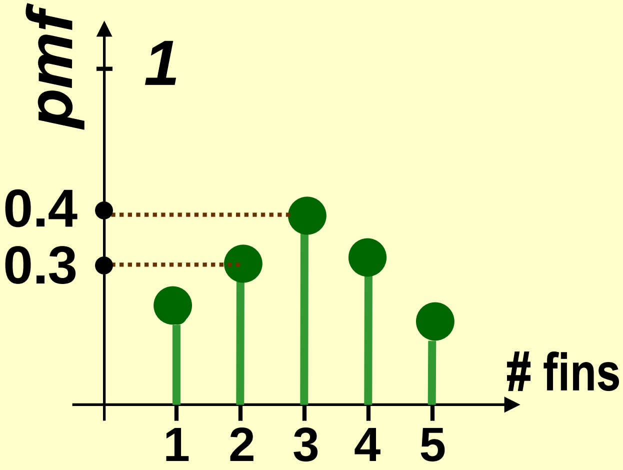
$$\frac{d}{dx}F(x)=f(x)$$

$$P(X=a)=\int_a^a f(x)dx=0$$

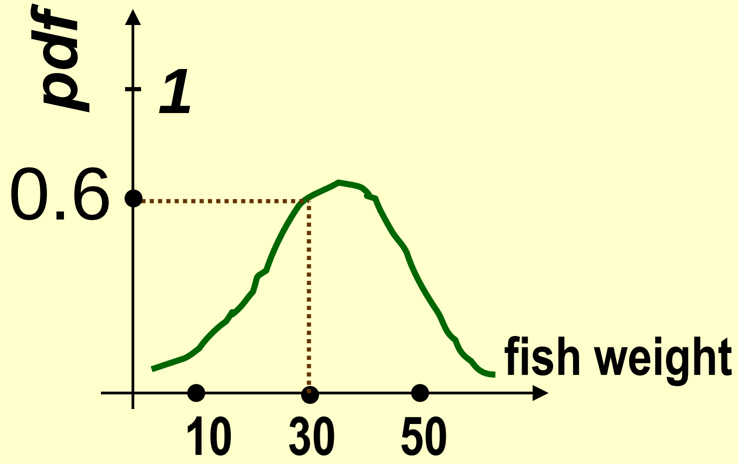
$$P(-\infty\le X\le\infty)=\int_{-\infty}^{\infty} f(x)dx=1$$

$$f(x)\ge 0$$

probability mass	probability density



- true probability
- P(fish has 2 or 3 fins)= =p(2)+p(3)=0.3+0.4
- take sums



- density, not probability
- P(fish weights 30kg) $\neq$ 0.6
- P(fish weights 30kg)=0
- P(fish weights between 29 and 31kg)= $\int_{29}^{31} f(x)dx$
- integrate

Expected Value

- Useful characterization of a r.v.
- Also known as mean, expectation, or first moment

discrete case: $\mu = E(X) = \sum_{\forall x} x \; p(x)$

continuous case: $\mu = E(X) = \int_{-\infty}^{\infty} x \, f(x) dx$

- Expectation can be thought of as the average over many experiments

Expected Value for Functions of X

- Let g(x) be a function of the r.v. X. Then

discrete case: $E[g(X)] = \sum_{\forall x} g(x)\, p(x)$

continuous case: $E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

- An important function of X: $[X-E(X)]^2$
  - Variance $E[[X-E(X)]^2] = var(X) = \sigma^2$
  - Variance measures the spread around the mean
  - Standard deviation = $[var(X)]^{1/2}$, has the same units as the r.v. X

Properties of Expectation

- If X is constant r.v. X=c, then E(X) = c

- If a and b are constants, E(aX+b)=aE(X)+b

- More generally,

$$E\left(\sum_{i=1}^{n}(a_i X_i + c_i)\right) = \sum_{i=1}^{n}(a_i E(X_i) + c_i)$$

- If a and b are constants, then
var(aX+b)= a^2var(X)

Pairs of Random Variables

- Say we have 2 random variables:
  - Fish weight ***X***
  - Fish lightness ***Y***
- Can define *joint* CDF

$$F(a,b) = P(X \le a, Y \le b) = P(\omega \in S \mid X(\omega) \le a, Y(\omega) \le b)$$

- Similar to single variable case, can define
  - discrete: joint probability mass function

$$p(a,b) = P(X = a, Y = b)$$

  - continuous: joint density function $f(x,y)$

$$P(a \le X \le b, c \le Y \le d) = \iint_{\substack{a \le x \le b \\ c \le y \le d}} f(x,y)\,dx\,dy$$

Marginal Distributions

- given joint mass function $p_{X,Y}(x,y)$, marginal, i.e. probability mass function for r.v. X can be obtained from $p_{X,Y}(x,y)$

$$p_X(x) = \sum_{\forall y} p_{X,Y}(x,y)$$

$$p_Y(y) = \sum_{\forall x} p_{X,Y}(x,y)$$

- marginal densities $f_X(x)$ and $f_Y(y)$ are obtained from joint density $f_{X,Y}(x,y)$ by integrating

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

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

Independence of Random Variables

- r.v. X and Y are independent if

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

- *Theorem*: r.v. X and Y are independent if and only if

$$p_{x,y}(x,y) = p_y(y)p_x(x)$$ ***(discrete)***

$$f_{x,y}(x,y) = f_y(y)f_x(x)$$ ***(continuous)***

More on Independent RV's

- If X and Y are independent, then
  - E(XY)=E(X)E(Y)
  - Var(X+Y)=Var(X)+Var(Y)
  - G(X) and H(Y) are independent

Covariance

- Given r.v. X and Y, covariance is defined as:

$$\text{cov}(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

- Covariance is useful for checking if features *X* and *Y* give similar information
- Covariance (from co-vary) indicates tendency of X and Y to vary together
  - If X and Y tend to increase together, Cov(X,Y) > 0
  - If X tends to decrease when Y increases, Cov(X,Y) < 0
  - If decrease (increase) in X does not predict behavior of Y, Cov(X,Y) is close to 0

Covariance Correlation

- If cov(X,Y) = 0, then X and Y are said to be uncorrelated (think unrelated). However X and Y are not necessarily independent.
- If X and Y are independent, cov(X,Y) = 0
- Can normalize covariance to get correlation

$$-1 \le cor(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} \le 1$$

Random Vectors

- Generalize from pairs of r.v. to vector of r.v. X= [X_1 X_2… X_3] (think multiple features)
- Joint CDF, PDF, PMF are defined similarly to the case of pair of r.v.'s

Example:

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

- All the properties of expectation, variance, covariance transfer with suitable modifications

Covariance Matrix

- characteristics summary of random vector
- $\mathrm{cov}(X)=\mathrm{cov}[X_1\ X_2 \ldots\ X_n] = \Sigma = E[(X-\mu)(X-\mu)^T]=$

$$\begin{bmatrix} E(X_1-\mu_1)(X_1-\mu_1) & \cdots & E(X_n-\mu_n)(X_1-\mu_1) \\ E(X_2-\mu_2)(X_1-\mu_1) & \cdots & E(X_n-\mu_n)(X_2-\mu_2) \\ \vdots & & \vdots \\ E(X_n-\mu_n)(X_1-\mu_1) & \cdots & E(X_n-\mu_n)(X_n-\mu_n) \end{bmatrix}$$

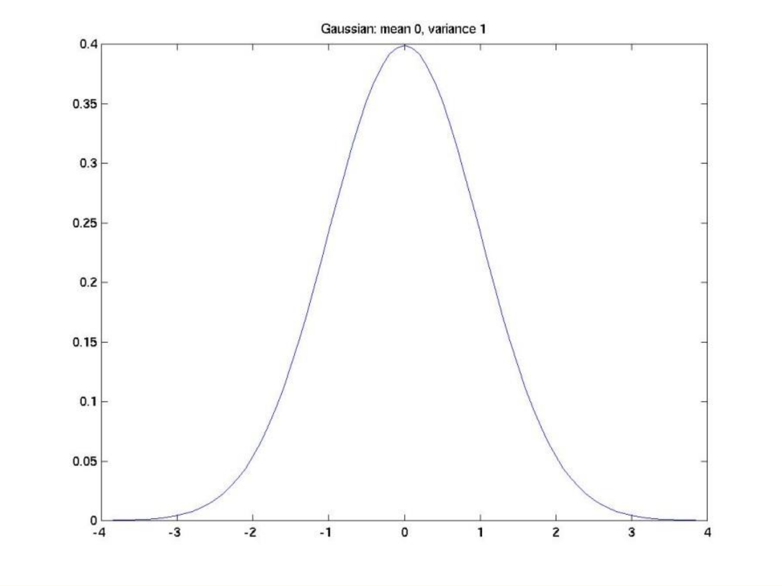
$$\begin{bmatrix} \sigma_1^2 & c_{12} & c_{13} \\ c_{21} & \sigma_2^2 & c_{23} \\ c_{31} & c_{32} & \sigma_3^2 \end{bmatrix}$$

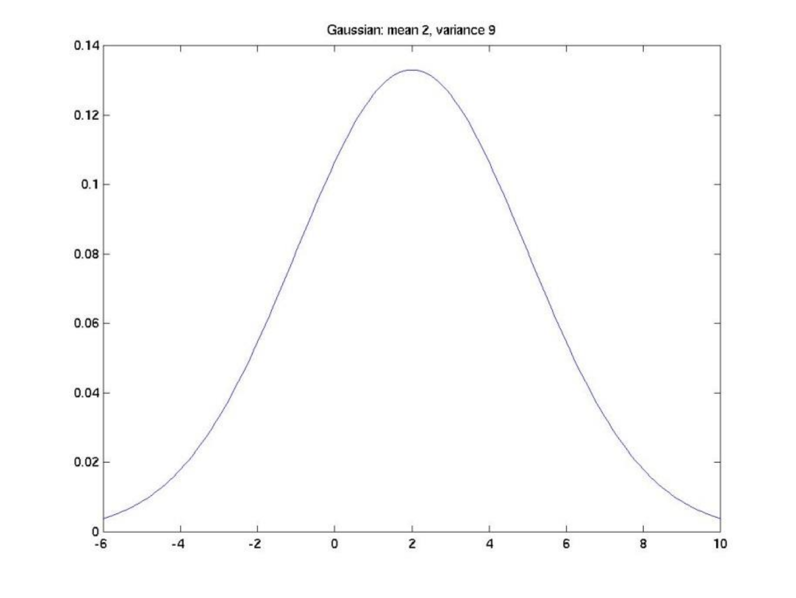
variances ***covariances***

Normal or Gaussian Random Variable

- Has density $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

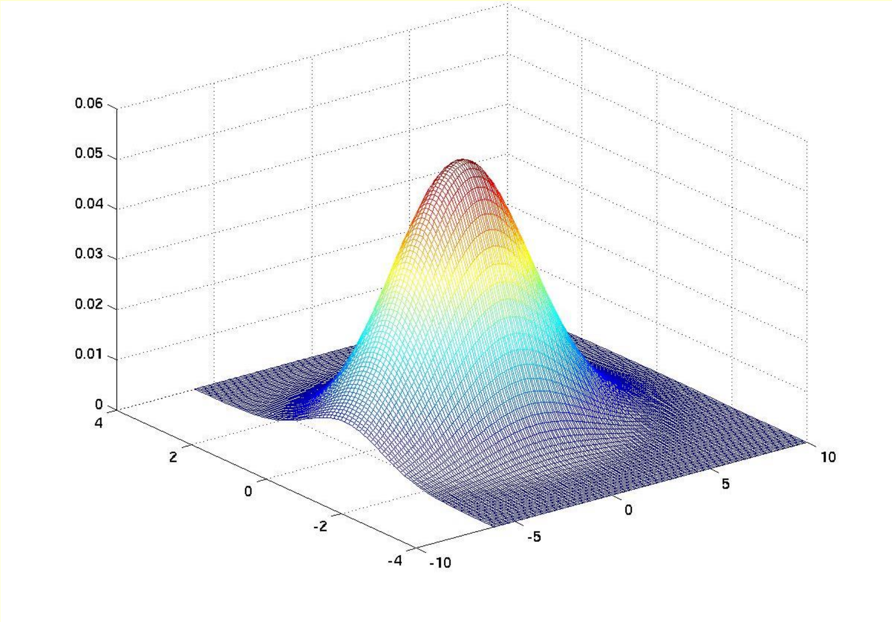
- Mean μ, and variance σ^2





Multivariate Gaussian

- has density $f(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}\left[(x-\mu)^t \Sigma^{-1} (x-\mu)\right]}$
- mean vector $\mu = \left[\mu_1, \ldots, \mu_n\right]$
- covariance matrix Σ



Conditional Mass Function, Bayes Rule

- Define conditional mass function of ***X*** given ***Y=y*** by

$$P(x \mid y) = \frac{P(x,y)}{P(y)}$$

y is fixed

- The law of Total Probability:

$$P(x) = \sum_{\forall y} P(x,y) = \sum_{\forall y} P(x \mid y) P(y)$$

- The Bayes Rule:

$$P(y \mid x) = \frac{P(y,x)}{P(x)} = \frac{P(x \mid y) P(y)}{\sum_{\forall y} P(x \mid y) P(y)}$$

Conditional Density Function, Bayes Rule

- Define conditional density function of ***X*** given ***Y***=***y*** by

$$p(x \mid y) = \frac{p(x,y)}{p(y)}$$

y is fixed

- The law of Total Probability:

$$p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy = \int_{-\infty}^{\infty} p(x \mid y)\,p(y)\,dy$$

- The Bayes Rule:

$$p(y \mid x) = \frac{p(y,x)}{p(x)} = \frac{p(x \mid y)\,p(y)}{\int_{-\infty}^{\infty} p(x \mid y)\,p(y)\,dy}$$