Vision and Image Processing: Recalls on Probabilities and Statistcs

François Lauze
(with a lot of inspiration from L. Sørensen and A. Feragen)

Department of Computer Science
University of Copenhagen

Plan for today

- Statistics. Observations, Means, Empirical Variance.
- Random Variables.
- Conditional Probabilities and Independence, Bayes Theorem
- Expectation, Variance, Moments.
- A Few Laws.

Introduction

- In real life, measurement process is not certain.
- Repeating measurements often leads to small (or not that small) variations on results.
- Some "average" frequencies / range of measurements appear in general more often that others.
- Probabilities: quantification of uncertainty.
- Random variables: modelling of the measurements.
- Statistics: the practical / empirical observation part.

Outline

Coin Example: Categorical Data

- 20 coin flips:

1	2	3	4	5	6	7	8	9	10
t	t	t	h	t	t	h	h	t	t
11	12	13	14	15	16	17	18	19	20
h	h	h	t	t	t	t	h	h	t

- Fair coin?
- Frequency of heads: 8/20, tails: 12/20.
- Not 10/20 = 1/2, but not far away!
- Can I compute the mean result? Does it make sense?
- Here, categorical data: half tail, half head does not really make sense.

Height Example: Numerical Data

- 48 samples of heights of adult Scandinavian females (in cm).

1	2	3	4	5	6	7	8	9	10	11	12
1.85	1.69	1.83	1.82	1.92	1.73	1.61	1.75	1.65	1.82	1.60	1.71
13	14	15	16	17	18	19	20	21	22	23	24
1.75	1.90	1.78	1.73	1.78	1.76	1.71	1.65	1.85	1.81	1.89	1.88
25	26	27	28	29	30	31	32	33	34	35	36
1.91	1.92	1.60	1.69	1.64	1.72	1.73	1.84	1.92	1.72	1.60	1.64
37	38	39	40	41	42	43	44	45	46	47	48
1.65	1.55	1.63	1.63	1.63	1.63	1.65	1.79	1.65	1.75	1.67	1.73

- Mean height? spread? repartition? Data is numerical here, this make sense.
- Max, Min, Mean, variance (average squared variation to the mean):

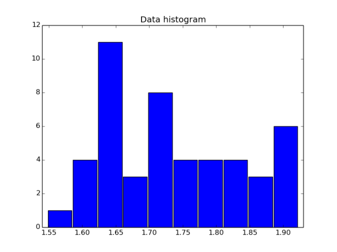
$$\max h_i = 1.92, \quad \min h_i = 1.55,$$

$$\bar{h} = \frac{1}{N}\sum_{i=1}^{N} h_i = 1.74, \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{n}\left(h_i - \bar{h}\right)^2 = 0.010$$

- Remember: standard deviation = $\sqrt{\text{variance}}$: $\sigma = 0.1$.

Representation

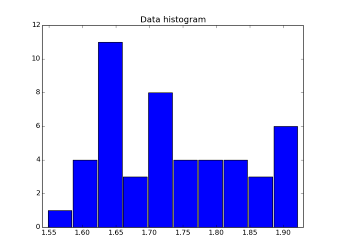
Range:	[1.55,1.58]	[1.59,1.62]	[1.63,1.66]	[1.67,1.69]	[1.70,1.73]	[1.74,1.77]	[1.78,1.80]	[1.81,1.84]	[1.85,1.88]	[1.89,1.92]
Count:	1	4	11	3	8	4	4	4	3	6



- Accumulate data in given ranges, a.k.a bins. 10 bins used here.
- Each bar displays amount of samples in a given range.
- Bin 3: range = [1.62, 1.66], 11 samples.
- Frequency that a sample is in range [1.63, 1.66]: 11/48.

Representation

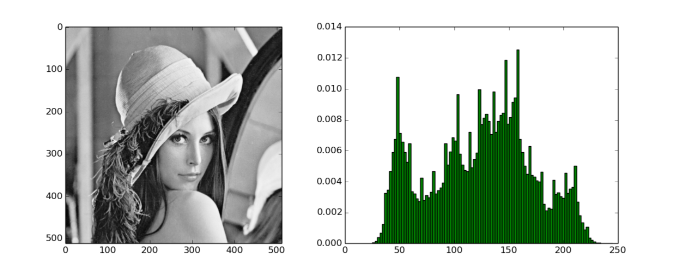
Range:	[1.55,1.58]	[1.59,1.62]	[1.63,1.66]	[1.67,1.69]	[1.70,1.73]	[1.74,1.77]	[1.78,1.80]	[1.81,1.84]	[1.85,1.88]	[1.89,1.92]
Count:	1	4	11	3	8	4	4	4	3	6



- Accumulate data in given ranges, a.k.a bins. 10 bins used here.
- Each bar displays amount of samples in a given range.
- Bin 3: range = [1.62, 1.66], 11 samples.
- Frequency that a sample is in range [1.63, 1.66]: 11/48.

- Tempting to say that “the probability that height is between 1.63 and 1.66 is 11/48”.
- With more samples maybe...

Image Histograms



- Each bin represents a range of grey level values (here 100 bins).
- Each bin value represents a frequency of these grey values in image.

Recap

- N outcomes of an experiment, results in given categories $c_1, \dots, c_k$.
- Can talk of frequency of apparition of category c_k:

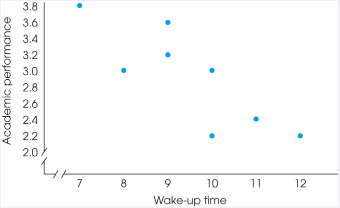
$$\frac{\text{number of outcomes in category } c_k}{\text{number of outcomes} = N}$$

- Can discuss of the largest represented category, smallest, mean category size...
- For numerical outcomes: grouping in ranges: histograms.
- Means, variances etc... make sense. Sample $h = (h_1, \dots, h_N)$:

$$\bar{h} = \frac{1}{N}\sum_{i=1}^{N} h_i, \quad \sigma_h^2 = \frac{1}{N}\sum_{i=1}^{N}\left(h_i - \bar{h}\right)^2 = \left(\frac{1}{N}\sum_{i=1}^{N} h_i^2\right) - \bar{h}^2$$

Bivariate Data

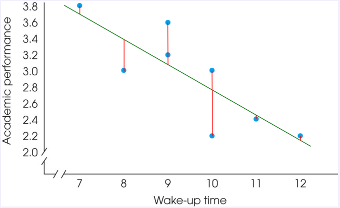
Child	Wake-up Time	Academic Performance
A	11	2.4
B	9	3.6
C	9	3.2
D	12	2.2
E	7	3.8
F	10	2.2
G	10	3.0
H	8	3.0



- Wake-up time and academic performances measured for a sample of 8 children. Correlated?

Bivariate Data

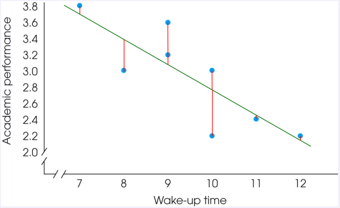
Child	Wake-up Time	Academic Performance
A	11	2.4
B	9	3.6
C	9	3.2
D	12	2.2
E	7	3.8
F	10	2.2
G	10	3.0
H	8	3.0



- Wake-up time and academic performances measured for a sample of 8 children. Correlated?
- It seems so, but the sample size is low!

Bivariate Data

Child	Wake-up Time	Academic Performance
A	11	2.4
B	9	3.6
C	9	3.2
D	12	2.2
E	7	3.8
F	10	2.2
G	10	3.0
H	8	3.0



- Wake-up time and academic performances measured for a sample of 8 children. Correlated?
- It seems so, but the sample size is low!
- And BTW, I got this data on the Net, no idea where it comes from, it could be fake!

Linear Regression, Covariance, Correlation Coefficient

$x = (x_1, \ldots, x_N)$ wake-up-time variable, $y = (y_1, \ldots, y_N)$ academic performance variable.

- Covariance between x and y

$$\mathrm{cov}(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y}$$

- Linear regression: Find a and b such that

$$\sum_{i=1}^{N} |a x_i + b - y_i|^2 = \min : a = \frac{\mathrm{cov}(x, y)}{\sigma_x^2}, \quad b \text{ is a bit too long}$$

  Here $a = -0.317$, $b = 5.933$.

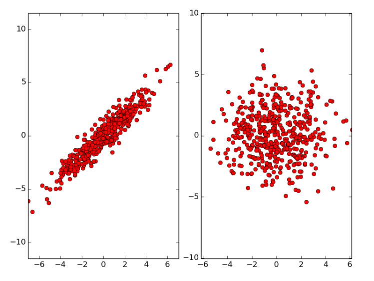
- How variations of one variable explains the other's: Pearson's correlation coefficient.

$$r = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y} = a\frac{\sigma_x}{\sigma_y} \in [-1, 1]$$

  Here $r = -0.828$.

Multivariate Data

Multivariate data $((x_1^1, x_2^1, \dots, x_p^1)^\top, (x_1^N, x_2^N, \dots, x_p^N)^\top$ with $\mathbf{x}^i = (x_1^i, x_2^i, \dots, x_p^i)^\top \in \mathbb{R}^p$.



- $p = 2$ here (else drawing complicated).
- Same spread along axes, but clear differences!
- Mean: $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^N \mathbf{x}^i \in \mathbb{R}^p$.
- Covariance Matrix:

$$\Sigma_{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^N (\mathbf{x}_i - \bar{x})(\mathbf{x}_i - \bar{x})^\top \in \mathbb{R}^{p\times p}$$

- In example from figure:

$$\text{Left: } \Sigma_{\mathbf{x}} = \begin{bmatrix} 3.80 & 3.45 \\ 3.45 & 3.61 \end{bmatrix}, \quad \text{right: } \Sigma_{\mathbf{x}} = \begin{bmatrix} 3.85 & 0.01 \\ 0.01 & 4.27 \end{bmatrix}$$

- In general: diagonal entry $(\Sigma_{\mathbf{x}})_{uu} = \sigma^2_{x_u}$, $(x_u = (x_u^1, \dots, x_u^N))$,
- off-diagonal entry $(\Sigma_{\mathbf{x}})_{uv} = \mathrm{cov}(x_u, x_v)$

Outline

A Discrete Case: Casting A Dice

- Here random variable: number returned after casting a dice X
- Observed value: x
- State space: Set of values of the random variable $S = \{1, 2, 3, 4, 5, 6\}$.
- Probability mass function

$$f(x) = \begin{cases} \frac{1}{6} & x \in S \\ 0 & \text{else.} \end{cases}$$

- Event: subset of S.

Events and Probabilities

- Event $E \subset S$.
- Examples:

$$E_1 = \text{odd} : E_1 = \{1, 3, 5\}$$
$$E_2 = \text{even} : E_2 = \{2, 4, 6\}$$
$$E_3 = X \leq 2 : E_3 = \{1, 2\}$$

- Probability of an event:

$$P(E_1) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$
$$P(E_2) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$
$$P(E_3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

- Dice casting: discrete random variable.

A Continuous Case: Weight of a Strawberries Basket

- Random Variable: weight of a basket.
- Observed weight x
- Probability density function: $f(x)$.
- State space $S = [0, \infty[$.
- Probability density function:

$$f(x) \geq 0, \quad \int_0^\infty f(x)\,dx = 1.$$

- Probability distribution:

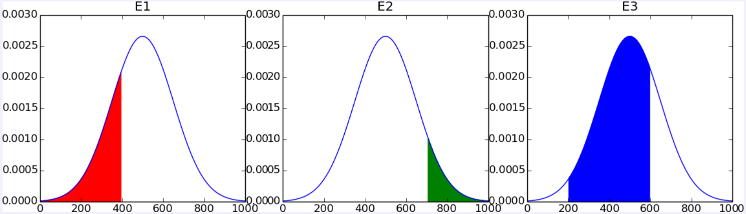
$$P(X < t) = \int_0^t f(x)\,dx$$

Weight of Strawberry Basket

- Event E_1: weight $X < 400$g. $P(E_1) = P(X < 400) = \int_0^{400} f(x)\,dx$.

- Event E_2: weight $X > 700$g. $P(E_2) = P(X > 700) = \int_{400}^{\infty} f(x)\,dx$.

- E_3: 200g $< X <$ 600g. $P(E_3) = P(200 < X < 600) = \int_{200}^{600} f(x)dx$



Conditional Probabilities

- An event probability may depend on another one!
- Probability that strawberry basket weight $<$ 500g in a supermarket depends on the probability that a customer ate one of them!
- Conditional Probability:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) > 0.$$

Conditional Probabilities

- An event probability may depend on another one!
- Probability that strawberry basket weight < 500g in a supermarket depends on the probability that a customer ate one of them!
- Conditional Probability:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) > 0.$$

  - $P(E|F)$: Probability of E knowing F.

Conditional Probabilities

- An event probability may depend on another one!
- Probability that strawberry basket weight < 500g in a supermarket depends on the probability that a customer ate one of them!
- Conditional Probability:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) > 0.$$

  - $P(E|F)$: Probability of E knowing F.
  - $P(E \cap F)$ common probability of E and F, i.e., probability that E and F are realized simultaneously.

Conditional Probabilities

- An event probability may depend on another one!
- Probability that strawberry basket weight < 500g in a supermarket depends on the probability that a customer ate one of them!
- Conditional Probability:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) > 0.$$

  - $P(E|F)$: Probability of E knowing F.
  - $P(E \cap F)$ common probability of E and F, i.e., probability that E and F are realized simultaneously.
  - Rewritten as Product Rule $P(E \cap F) = P(E|F)P(F)$

Conditional Probabilities

- An event probability may depend on another one!
- Probability that strawberry basket weight < 500g in a supermarket depends on the probability that a customer ate one of them!
- Conditional Probability:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) > 0.$$

  - $P(E|F)$: Probability of E knowing F.
  - $P(E \cap F)$ common probability of E and F, i.e., probability that E and F are realized simultaneously.
  - Rewritten as Product Rule $P(E \cap F) = P(E|F)P(F)$

Conditional Probabilities

- An event probability may depend on another one!
- Probability that strawberry basket weight < 500g in a supermarket depends on the probability that a customer ate one of them!
- Conditional Probability:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) > 0.$$

  - $P(E|F)$: Probability of E knowing F.
  - $P(E \cap F)$ common probability of E and F, i.e., probability that E and F are realized simultaneously.
  - Rewritten as Product Rule $P(E \cap F) = P(E|F)P(F)$

- Total Probability Theorem, important in classification! $(E_i)_i$ a pairwise complete disjoint set of events $E_i \cap E_j = \emptyset$, $i \neq j$, $\cup_i E_i = S$. Then

$$P(F) = P(F|E_1)P(E_1) + P(F|E_2)P(E_2) + \cdots + P(F|E_n)P(E_n)$$

Dice Exercise

- Two dice.
- Random Variables X: dice 1 value, and Y: dice 2 value.
- Events:

$$\begin{cases} E_1 : & X = 1 \\ E_2 : & X + Y = 4 \\ E_3 : & Y = 4 \end{cases}$$

- Probabilities of Events:

$$P(E_1) = \frac{1}{6}, \quad P(E_2) = \frac{3}{36}, \quad P(E_3) = \frac{1}{6}.$$

- Combined Probabilities:

$$P(E_1 \cap E_2) = \frac{1}{36}, \quad P(E_1 \cap E_3) = \frac{1}{36}, \quad P(E_2 \cap E_3) = 0.$$

Dice Exercise

- Cast the two dice. Dice 1 random variable is Y. Observed: Dice 1 (X) shows 1 ($X = 1$).
  Probability that the sum of dice is 3, i.e. $X + Y = 3$ (event E_2).

Dice Exercise

- Cast the two dice. Dice 1 random variable is Y. Observed: Dice 1 (X) shows 1 ($X = 1$).
  Probability that the sum of dice is 3, i.e. $X + Y = 3$ (event E_2).

- Apply Conditional Probability Formula.

Dice Exercise

- Cast the two dice. Dice 1 random variable is Y. Observed: Dice 1 (X) shows 1 ($X = 1$).
  Probability that the sum of dice is 3, i.e. $X + Y = 3$ (event E_2).

- Apply Conditional Probability Formula.

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{\frac{1}{36}}{6} = \frac{6}{36} = \frac{1}{6}$$

Statistical Independence

- Two events E and F are independent if the following equivalent conditions are satisfied

$$P(E|F) = P(E) \iff P(E \cap F) = P(E)P(F).$$

- Back to the Dice: We observe Dice 2: Probability for $Y = 4$ (event E_3):

$$P(E_3|E_1) = \frac{\frac{1}{36}}{6} = \frac{1}{6} = P(E_3).$$

  Knowledge of event E_1 does not influence event E_3.

Bayes Theorem

- Many forms, very useful, many interpretations!
- Assume some statistical knowledge of event E is given as $P(E)$.
- Then assume a new event F is realized, we have $P(F|E)$.
- Bayes Theorem allows to update knowledge on E as $P(E|F)$,

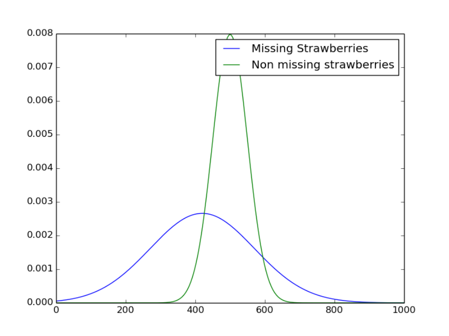
$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$

- $P(E)$ is called the prior (knowledge).
- $P(E|F)$ is called the posterior.
- $P(F|E)$ is called the likelihood.
- $P(F)$ is called the evidence.
- Proof is straightforward.

$$P(E|F) := \frac{P(E \cap F)}{P(F)} = \frac{P(E \cap F)}{P(E)} \frac{P(E)}{P(F)} = \frac{P(F|E)P(E)}{P(F)}$$

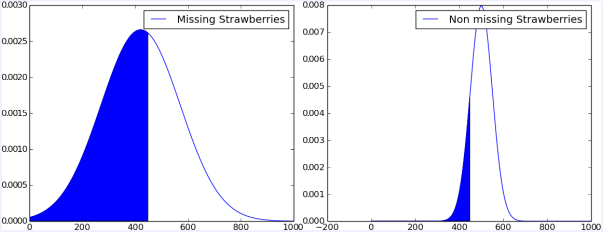
Strawberries Again

- Events.
  - Event E: someone ate some of my strawberries.
  - Event F: $X < 450$.g
- Prior knowledge.
  - P(someone ate some of my strawberries) $= 0.75$ (I am suspicious by nature).
  - P(no one ate my strawberries) $= 0.25$).



- I have a basket with less than 450g strawberries.

$$P(X < 450|\text{ missing}) = 0.58$$
$$P(X < 450|\text{ not missing}) = 0.16 \text{ (risky business)}$$



- Update our knowledge thanks to Bayes:

$$P(\text{eaten } |X < 450) = \frac{P(X < 450|\text{ eaten})P(\text{eaten})}{P(X < 450)}$$
$$= \frac{0.58 * 0.75}{0.58 * 0.75 + 0.16 * 0.25} = 0.92$$

Moments of a Discrete Random Variable

- Discrete variable X, state space $\{x_1, x_2, \dots x_k\} \subset \mathbb{R}$ with probability mass $P(X)$: $P(x_i) = p_1$ with $p_1 + p_2 + \dots p_k = 1$, $p_i >= 0$.
- Expectation of X:

$$E(X) = x_1 p_1 + x_2 p_2 + \dots x_k p_k = \sum_{i=1}^{k} x_i p_i = \sum_{i=1}^{k} x_i P(x_i)$$

- In general, if $F : \mathbb{R} \to \mathbb{R}$ a function which transform the value of X, $F(X)$ new random variable with expectation

$$E(F(X)) = \sum_{i=1}^{k} f(x_i) p_i$$

- Variance of X:

$$\sigma_X = \sum_{i=1} (n_i - E(X))^2 p_i = E\left((X - E(X))^2\right) = E(X^2) - E(X)^2$$

- Order n moments, non centered and centered:

$$E(X^n) = \sum_{i=1}^{n} x_i^n p_i, \quad E\,(X - E(X))^n = \sum_{i=1}^{n} (x_i - E(X))^n p_i$$

Moments of a Continuous Random Variable

- Continuous variable X, state space $S = \mathbb{R}$ with probability density function $p(x)$,
$$\int_{-\infty}^{\infty} p(x)\,dx = 1.$$
- Expectation of X:
$$E(X) = \int_{-\infty}^{\infty} x\,p(x)\,dx$$
- In general, if $F : \mathbb{R} \to \mathbb{R}$ a function which transform the value of X, $F(X)$ new random variable with expectation
$$E(F(X)) = \int_{-\infty}^{\infty} F(x)\,p(x)\,dx$$
- Variance of X:
$$\sigma_X = \int_{-\infty}^{\infty} (x - E(x))^2\;p(x)\,dx = E\left(X - E(X)\right)^2 = E(X^2) - E(X)^2.$$
- Order n moments, non centered and centered:
$$E(X^n) = \int_{-\infty}^{\infty} x^n\,dx, \quad E\left(X - E(X)\right)^n = \int_{-\infty}^{\infty} (x - E(x))^n\;dx.$$

Discrete Vector Valued Random Variable

- item $\mathbf{X} = (X, Y)^\top$ position on a grid. State space: a finite subset of $\mathbb{R}^n$, probability mass $P(X, Y)$ with

$$p(x_i, y_i) \geq 0, \quad \sum_{i=1}^{n} p(x_i, y_i) = 1.$$

- Expectation Vector: $E(\mathbf{X})$: component-wise expectation/average

$$E(\mathbf{X}) = E((X, Y)^\top) = \sum_{i=1}^{n} \begin{bmatrix} x_i \\ y_i \end{bmatrix} p(x_i, y_i) = \begin{bmatrix} \sum_{i=1}^{n} x_i p(x_i, y_i) \\ \sum_{i=1}^{n} y_i p(x_i, y_i) \end{bmatrix} = \begin{bmatrix} E(X) \\ E(Y) \end{bmatrix}$$

- Covariance Matrix $\Sigma_{\mathbf{X}}$

$$\Sigma_{\mathbf{X}} = \sum_{i=1}^{n} \underbrace{\left(\begin{bmatrix} x_i \\ y_i \end{bmatrix} - E(\mathbf{X}) \right) \left(\begin{bmatrix} x_i & y_i \end{bmatrix} - E(\mathbf{X})^\top \right)}_{\text{outer product}} = \sum_{i=1}^{n} \begin{bmatrix} x_i - E(X) \\ y_i - E(Y) \end{bmatrix} \left[x_i - E(X) \right.$$

$$= \sum_{i=1}^{n} \begin{bmatrix} (x_i - E(X))^2 & (x_i - E(X))(y_i - E(Y)) \\ (x_i - E(X))(y_i - E(Y)) & (y_i - E(Y))^2 \end{bmatrix}$$

Continuous Vector-valued Random Variables

- $\mathbf{X} = (X, Y)^\top$ position of a bullet/dart on a target $(x, y)^\top$. State space $S = \mathbb{R}^2$ (well...). Probability density function $p(x, y) \geq 0$,

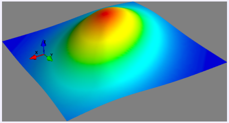
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)\, dx, dy = 1.$$

- Probability distribution

$$P(X < s, Y < t) = \int_{-\infty}^{s}\int_{-\infty}^{t} f(x, y)\, dx\, dy.$$

- Expectation vector

$$E(\mathbf{X}) = (E(X), E(Y))^\top = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x, y)^\top p(x, y)\, dxdy$$



Outline

Bernoulli Distribution

- A simple random variable X with two states: $S = \{0, 1\}$
- Probability mass function $P((X = 0) = p \in]0, 1[$, $P(X = 1) = 1 - p$.
- p is the parameter of the distribution.
- Expectation:

$$E(X) = 0.p + 1.(1 - p) = 1 - p$$

- Variance:

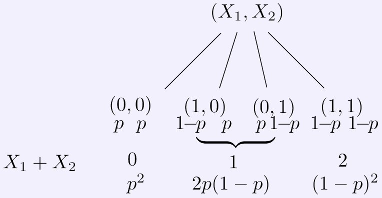
$$\sigma_X^2 = E(X-(1-p))^2 = (0-(1-p))^2.p+(1-p-(1-p))^2.(1-p) = p(1-p)^2$$

- Spread/standard deviation = $\sqrt{\text{variance}}$:

$$\sigma_X = (1 - p)\sqrt{p}$$

Binomial Distribution

- Number of success (1) in n repeated Bernoulli experiment.
- $X_1, \dots, X_n$ random variables each with Bernoulli distribution with same parameter p.
- Binomial distribution with parameters (p, n): $X = X_1 + X_2 \cdots + X_n$. State space: $S = \{0, 1, \dots, n\}$.
- Probability mass?

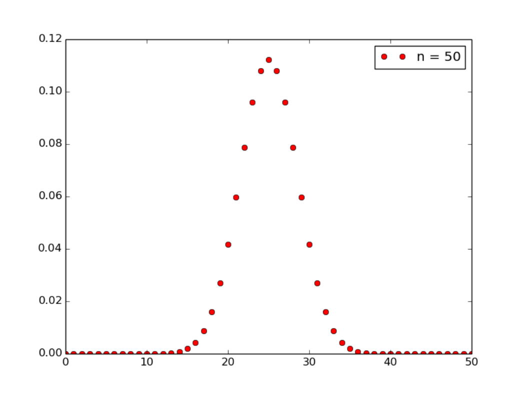


$$P(X=0) = p^2, \quad P(X=1) = 2p(1-p), \quad P(X=2) = (1-p)^2.$$

- Binomial formula for $n = 2$:
  $1 = 1^2 = (p + (1-p))^2 = p^2 + 2p(1-p) + (1-p)^2$!
- Expectation, variance?

Binomial Law: Excercise

- $n = 3$, $X = X_1 + X_2 + X_3$. How many possible cases for values of triple (X_1, X_2, X_3)
- Probability mass for X
- Moments (expectation, variance...).



Uniform Distribution on a Segment

- Continuous. State space $S = [a, b] \in \mathbb{R}$. X random variable: all values are equiprobable.
- Probability density function

$$p(x) = \frac{1}{b-a}, \quad \text{independent of } x! : \int_a^b \frac{dx}{b-a} = 1$$

- Expectation:

$$E(X) = \int_a^b \frac{x}{b-a}\, dx = \frac{a+b}{2}$$

  Average position is the mid-point!
- Variance:

$$\sigma_X^2 = \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2} \right)^2 dx = \frac{(b-a)^2}{12}$$

1D Gaussian

- Continuous Random Variable X. Maybe the most important one, due to the celebrated Central Limit Theorem.
- State space $\mathbb{R}$.
- Parameters: $\mu \in \mathbb{R}$, $\sigma \in]0, \infty[$.
- Probability density function

$$p_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Classical theorem shows that

$$\int_{-\infty}^{\infty} p_{\mu,\sigma}(x)\, dx = 1.$$

- Expectation: $E(X) = \mu$.
- Variance $\sigma_X^2 = \sigma$.
- Continuous limiting case for normalized binomial distribution when $n \to \infty$.

The End

Please read Forsyth and Ponce chapter on probabilities.