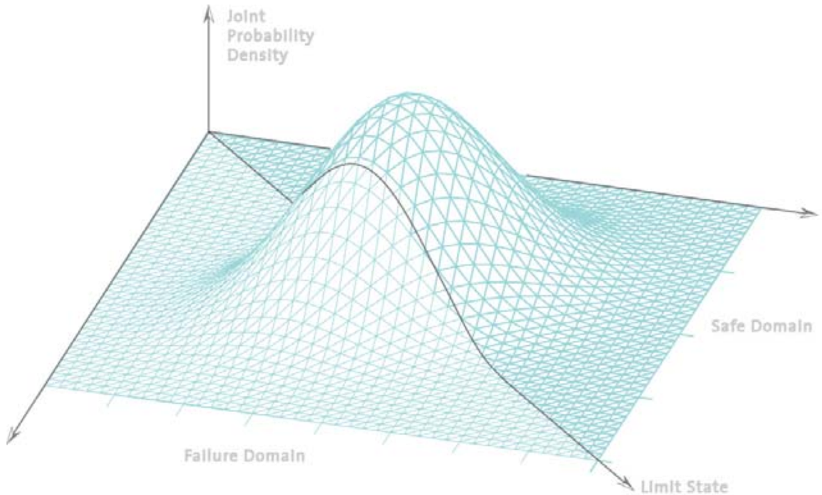


Exercises Tutorial 6

Statistics and Probability Theory

Prof. Dr. Michael Havbro Faber
Swiss Federal Institute of Technology Zurich
ETHZ

Assessment 1:

Thursday **3rd of May**

Where?: **HCI G7**

Time?: **8.00-** try to be exact on your time!

Allowed?: Everything (script and the course's material is more than enough)
Not Allowed? : mobile phones, pc's etc that provide communication means

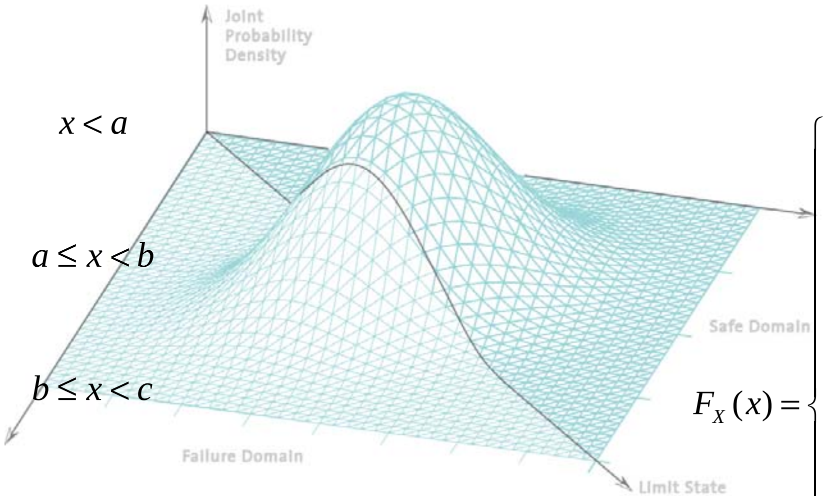
Up to which lecture? Lecture 6

Do not forget:

- Legi
- Calculator
- Pens!

Please correct in the solutions of the tutorial exercises:

$$F_X(x) = \begin{cases} 0 & x < a \\ h \cdot \dfrac{(x-a)^2}{2\cdot(b-a)} + C_1 & a \le x < b \\ h \cdot x + C_2 & b \le x < c \\ h \cdot \dfrac{(x-d)^2}{2\cdot(c-d)} + C_3 & c \le x < d \\ \cancel{0}\;1 & d \le x \end{cases}$$



$$F_X(x) = \begin{cases} 0 & x < 1 \\ \dfrac{(x-1)^2}{6} & 1 \le x < 2 \\ \dfrac{x}{3} - \dfrac{1}{2} & 2 \le x < 3 \\ -\dfrac{(x-6)^2}{18} + 1 & 3 \le x < 6 \\ \cancel{0}\;1 & 6 \le x \end{cases}$$

On pages 4.2 and 4.4

Correct also the numbering of exercise 6.2

Exercise 5.3 (Group exercise)

Highway bridges may require maintenance in their life time. The duration where no maintenance is required, T, is assumed exponentially distributed with the mean value of 10 years. The maintenance activity takes some time, which is represented by S. The time is also assumed exponentially distributed with the mean value of 1/12 year.

a. Assuming that T and S are independent, obtain the **distribution of the time** between which subsequent maintenance activities are initiated, Z, i.e., $\boldsymbol{Z=S+T}$.

b. How large is the probability $P(Z \leq 5)$?

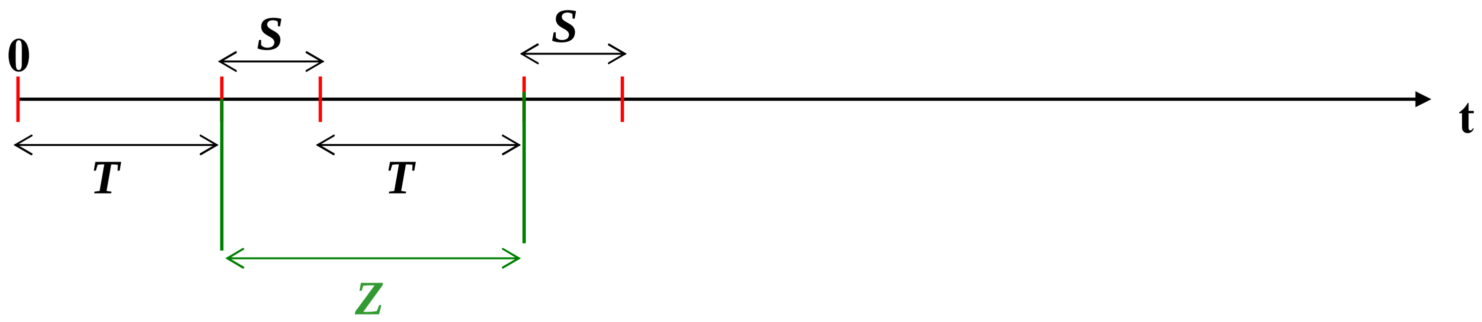
c. Assume that two bridges in a highway system are opened and the times until the bridges require the maintenance are represented by T_1 and T_2 which are independent identically distributed as T.
   How large is the probability that in the next 5 years no maintenance is required for the two bridges?

Exercise 5.3 (Group exercise)

Time where no maintenance is required, T: exponentially distributed $\mu_T = 10$ years

Time of maintenance S: exponentially distributed $\mu_S = \dfrac{1}{12}$ years

a. Assuming that T and S are independent, obtain the distribution of the time between subsequent maintenance activities are initiated, Z, i.e., $Z=S+T$.



In the case of continuous random variables: $Z=X+Y$

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

Exercise 5.2 a.

Measurements of wind speed are taken with an accurate and a less accurate device.

The joint probability of the measurements by both devices of the number of days in which the wind speed exceeds a threshold.

N_U represents the number of the days when the wind speed measured with the *accurate device* exceeds the threshold, and
N_G represents the number of the days when the wind speed measured with the *less accurate device* exceeds the threshold.

	$N_U = 0$	$N_U = 1$	$N_U = 2$	$N_U = 3$	$P(N_G)$
$N_G = 0$	0.2910	0.0600	0.0000	0.0000	**0.3510**
$N_G = 1$	0.0400	0.3580	0.0100	0.0000	0.4080
$N_G = 2$	0.0100	0.0250	0.1135	0.0300	0.1785
$N_G = 3$	0.0005	0.0015	0.0100	0.0505	0.0625
$P(N_U)$	0.3415	0.4445	0.1335	0.0805	$\sum = 1.00$

a. Calculate the probability that the number of days at which the wind speed, measured by each device, exceeds the threshold coincides.

$$P[N_U = N_G] = 0.2910 + 0.3580 + 0.1135 + 0.0505 = 0.813$$

Exercise 5.2 b.

b. Assume that the accurate device always measures the exact wind speed.
What are the probabilities that the wind speed which exceeds the threshold prevails 0, 1, 2 and 3 time(s) in a year when the wind speed measured with the less accurate device exceeds the threshold twice?

Conditional probability-Baye's rule $P[N_U \mid N_G = 2] = \dfrac{P[N_U \cap (N_G = 2)]}{P[N_G = 2]}$

	$N_U = 0$	$N_U = 1$	$N_U = 2$	$N_U = 3$	$P(N_G)$
$N_G = 0$	0.2910	0.0600	0.0000	0.0000	**0.3510**
$N_G = 1$	0.0400	0.3580	0.0100	0.0000	0.4080
$N_G = 2$	0.0100	0.0250	0.1135	0.0300	0.1785
$N_G = 3$	0.0005	0.0015	0.0100	0.0505	0.0625
$P(N_U)$	0.3415	0.4445	0.1335	0.0805	$\sum = 1.00$

$$P[N_U = 0 \mid N_G = 2] = \frac{P[(N_U = 0) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.01}{0.1785} = 0.056$$

$$P[N_U = 1 \mid N_G = 2] = \frac{P[(N_U = 1) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.025}{0.1785} = 0.1401$$

$$P[N_U = 2 \mid N_G = 2] = \frac{P[(N_U = 2) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.1135}{0.1785} = 0.6359$$

$$P[N_U = 3 \mid N_G = 2] = \frac{P[(N_U = 3) \cap (N_G = 2)]}{P[N_G = 2]} = \frac{0.03}{0.1785} = 0.1681$$

Exercise 6.1

Let $\{X_i\}_{1\ \ i\ \ 50}$ be independent, identically Normal distributed with mean value of $\mu = 1$ and standard deviation of $\sigma = 2$. Define:

$$S_n = X_1 + X_2 + \ldots + X_n \qquad \text{and} \qquad \bar{X}_n = \frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right) = \frac{S_n}{n}$$

where n=50.

a) Calculate the mean and the standard deviation of S_n and $\bar{X}_n$.

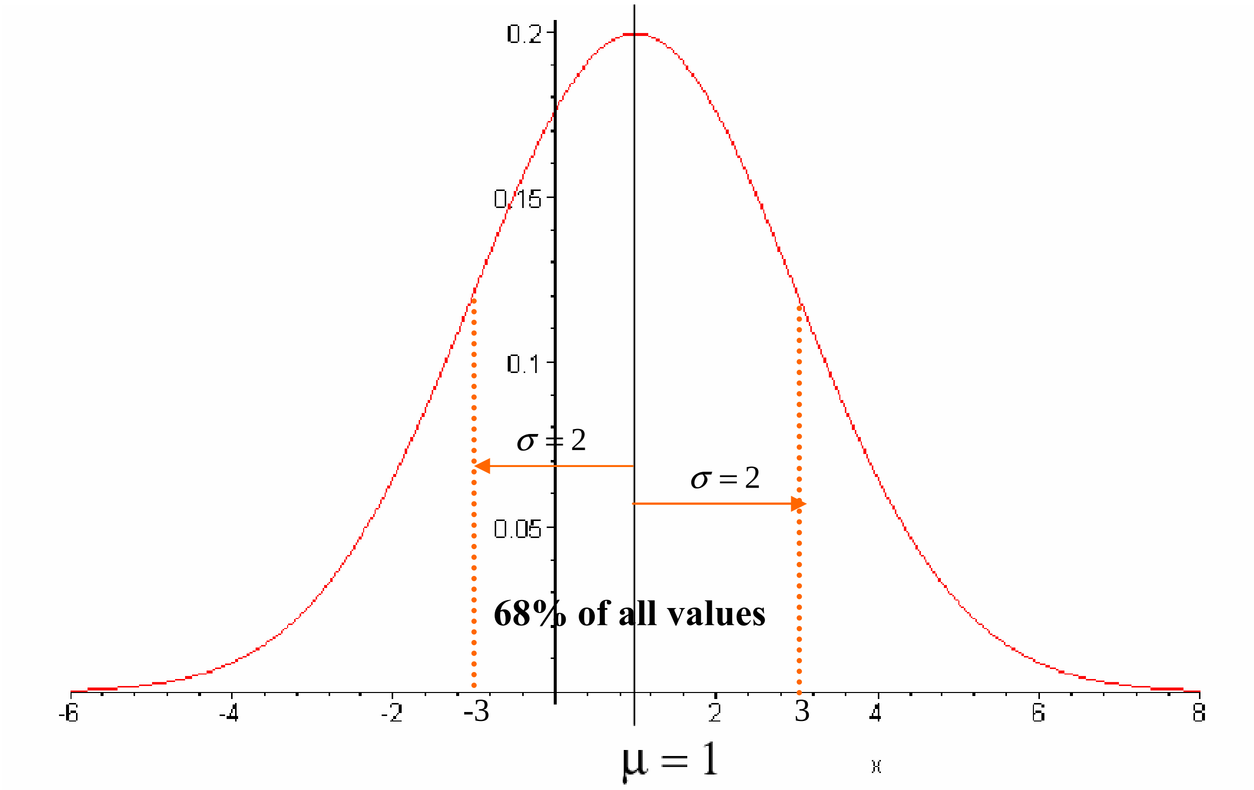
b) Calculate $P\left(E\left[X_1\right]-1 \le X_1 \le E\left[X_1\right]+1\right)$

c) Calculate $P\left(E\left[S_n\right]-1 \le S_n \le E\left[S_n\right]+1\right)$

d) Calculate $P\left(E\left[\bar{X}_n\right]-1 \le \bar{X}_n \le E\left[\bar{X}_n\right]+1\right)$

Exercise 6.1

Probability density function of Normal distribution with mean value = 1 and standard deviation = 2



Exercise 6.1

a) Calculate the mean and the standard deviation of S_n.

$$S_n = X_1 + X_2 + \ldots + X_n$$

Sum of Normal random variables is also a Normal distributed random variable!

Remember the following formulas:

$$E[X+Y] = E[X] + E[Y]$$
$$Var[X+Y] = Var[X] + Var[Y]$$

$$\mu_{S_n} = E[S_n] = \boxed{}$$

$$\sigma_{S_n}{}^2 = Var[S_n] = \boxed{}$$

$$N(\mu_{S_n}, \sigma_{S_n}{}^2) = N(\ldots\ldots, \ldots\ldots)$$

Exercise 6.1

a) Calculate the mean and the standard deviation of $\overline{X}_n$.

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) = \frac{S_n}{n}$$

Remember the following formulas:

$$E[cX] = cE[X]$$

$$Var[cX] = c^2 Var[X]$$

$$E\left[\overline{X}_n\right] = \boxed{}$$

$$V\left[\overline{X}_n\right] = \boxed{}$$

$$N(\mu_{\overline{X}_n}, \sigma_{\overline{X}_n}{}^2) = N(\ldots\ldots,\ldots\ldots)$$

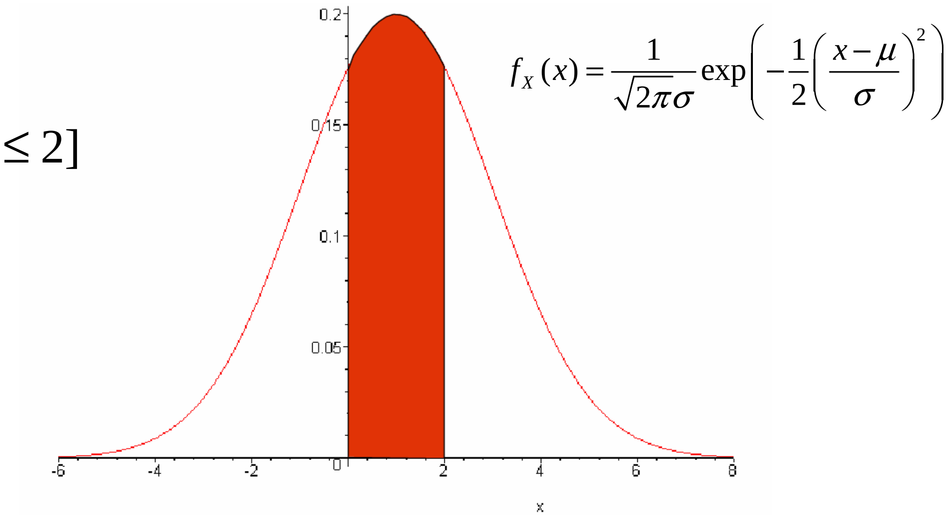
Exercise 6.1

b) Calculate $P\left(E\left[X_1\right]-1 \leq X_1 \leq E\left[X_1\right]+1\right)$

$$P\left(E\left[X_1\right]-1 \leq X_1 \leq E\left[X_1\right]+1\right) = P[0 \leq X_1 \leq 2]$$

$$= \boxed{\int_0^2 \frac{1}{\sqrt{2\pi}\cdot 2} \exp\left(-\frac{(x-1)^2}{2\cdot 2^2}\right) dx}$$

Use table in the script page D-17



What is a probability table?

How to integrate??? -> Utilize the probability table!

In order to utilize the probability table, the variable has to be standardized!

Exercise 6.1

Standardization

$$P\left(E[X_1]-1 \le X_1 \le E[X_1]+1\right) = P[0 \le X_1 \le 2] \qquad Z = \frac{X_1 - \mu}{\sigma}$$

$$= P[0-1 \le X_1 - 1 \le 2-1]$$ ← Mean value of $\mu = 1$

$$= P[\frac{0-1}{2} \le \frac{X_1 - 1}{2} \le \frac{2-1}{2}]$$ ← Standard deviation of $\sigma = 2$

$$= P[-\frac{1}{2} \le \frac{X_1 - 1}{2} \le \frac{1}{2}]$$

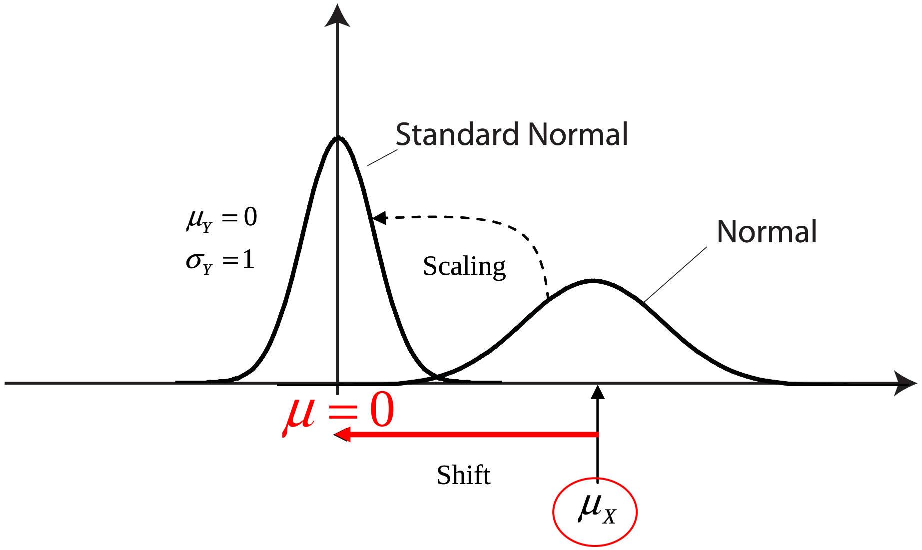
$$= P[-\frac{1}{2} \le Z \le \frac{1}{2}]$$

$$= \Phi(\frac{1}{2}) - \Phi(-\frac{1}{2})$$

$\Phi(z)$ is the cumulative distribution function for the Standard Normal distributed random variable N(0,1^2)

Exercise 6.1

Standardization



$$Z = \frac{X_1 - \mu}{\sigma}$$

$$E[Z] = E[\frac{X_1 - \mu}{\sigma}] = \frac{E[X_1] - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

$$Var[Z] = Var[\frac{X_1 - \mu}{\sigma}] = \frac{1}{\sigma^2} Var[X_1 - \mu]$$

$$= \frac{1}{\sigma^2} Var[X_1] = \frac{\sigma^2}{\sigma^2} = 1$$

$\Phi(z)$ is the cumulative distribution function for the Standard Normal distributed random variable $N(0,1^2)$

Exercise 6.1

Probability table

$$P\left(E\left[X_1\right]-1 \leq X_1 \leq E\left[X_1\right]+1\right)$$
$$=\Phi\left(\frac{1}{2}\right)-\Phi\left(-\frac{1}{2}\right)$$

Where is Φ(-0.5)?? Nowhere in the table!

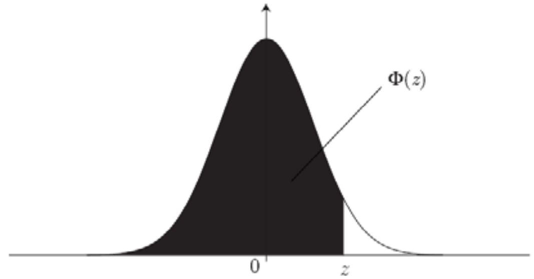
because… $\Phi(-z)=1-\Phi(z)$

so that $\Phi(-0.5)=1-\Phi(0.5)$

$$=1-0.6915$$
$$=0.3085$$

Finally

$$P\left(E\left[X_1\right]-1 \leq X_1 \leq E\left[X_1\right]+1\right)$$
$$=0.6915-0.3085=0.3830$$



Probability density function of the standard normal random variable.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.10	0.9821356
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.20	0.9860966
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.30	0.9892759
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.40	0.9918025
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.50	0.9937903
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.60	0.9953388
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.70	0.9965330
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.80	0.9974449
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.90	0.9981342
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	3.00	0.9986501
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	3.10	0.9990324
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	3.20	0.9993129
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	3.30	0.9995166
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	3.40	0.9996631
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	3.50	0.9997674
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	3.60	0.9998409
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	3.70	0.9998922
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	3.80	0.9999277
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	3.90	0.9999519
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	4.00	0.9999683
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	4.10	0.9999793

Table T.1 in Annex T of the Script

Exercise 6.1

c) Calculate $P\left(E\left[S_n\right]-1 \leq S_n \leq E\left[S_n\right]+1\right)$

→ same steps:

$$P\left(E\left[S_n\right]-1 \leq S_n \leq E\left[S_n\right]+1\right)$$

$$= P[\ldots\ldots\ldots \leq S_n \leq \ldots\ldots\ldots]$$

$$= P[\ldots\ldots\ldots\ldots \leq \ldots\ldots\ldots\ldots \leq \ldots\ldots\ldots\ldots]$$

$$= \Phi\left(\ldots\ldots\ldots\ldots\ldots\ldots\right) - \Phi\left(\ldots\ldots\ldots\ldots\ldots\ldots\right)$$

$$=$$

Find the standardized form $Z = \dfrac{X_1 - \mu}{\sigma}$

Look up the values in the probability table

Subtract $F_b - F_a$

Exercise 6.1

d) Calculate $P\left(E\left[\bar{X}_n\right]-1 \leq \bar{X}_n \leq E\left[\bar{X}_n\right]+1\right)$

→same steps:

-Standardization
-Probability table

$$P\left(E\left[\bar{X}_n\right]-1 \leq \bar{X}_n \leq E\left[\bar{X}_n\right]+1\right)$$

$$= P[.........\leq \bar{X}_n \leq ....]$$

$$= P[...........\leq ....\leq ....]$$

$$= \Phi\left(.................\right) - \Phi\left(..................\right)$$

$$=$$

Exercise 6.2

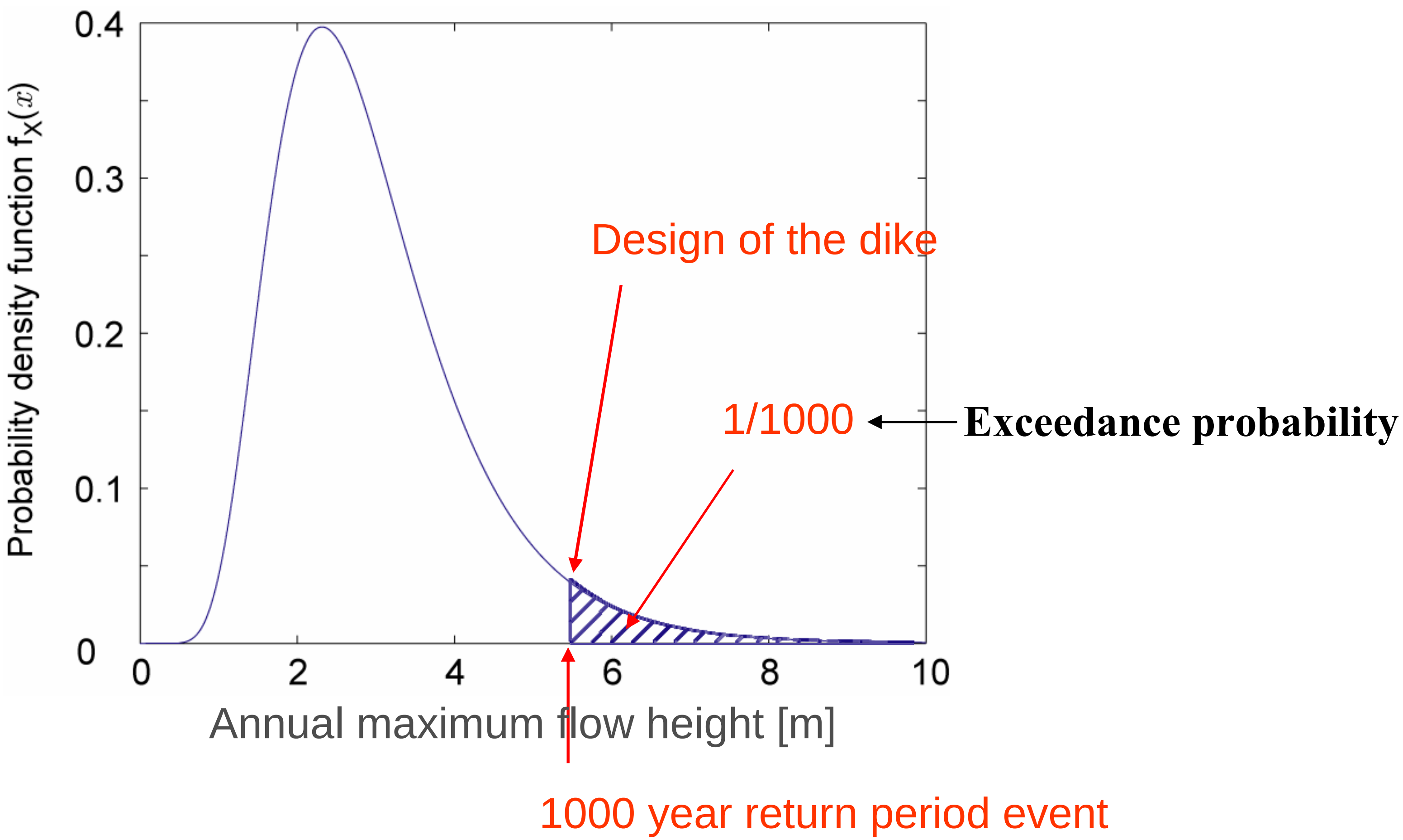
A dike is designed to withstand a “1000-year return period flood”. How large is the probability that water will overflow the dike in the following conditions:

a) During a 10 year period, for the first time in a given year?
b) During a 10 year period, twice?
c) Will not overflow during a 10 year period?
d) During a 10 year period, at most once?
e) During a 100 year period, 10 times?
f) -
g) During a 1000 year period, once or more often?

(It is assumed that flood occurs once in a year.)

Exercise 6.2

Return period T :



Exercise 6.2

Return period T:

Annual exceedance probability is p (= 1/T).

Random variable N = time until a flood occurs for the first time

The probability that a flood occurs in the n^{th} year is

$$
\begin{aligned}
P[N=n] &= \underbrace{(1-p)(1-p)\ldots(1-p)}_{n-1}\,p \\
&= (1-p)^{n-1}\,p
\end{aligned}
$$

← Geometric distribution

Expected value of N, $E[N]$ is

$$
E[N] = \sum_{n=1}^{\infty} nP[N=n] = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = \frac{1}{p} = T
$$

Exercise 6.2

The probability that a flood occurs in the n^{th} year is

$$P[N=n] = \underbrace{(1-p)(1-p)...(1-p)}_{n-1}\,p \qquad \longleftarrow \text{Geometric distribution}$$

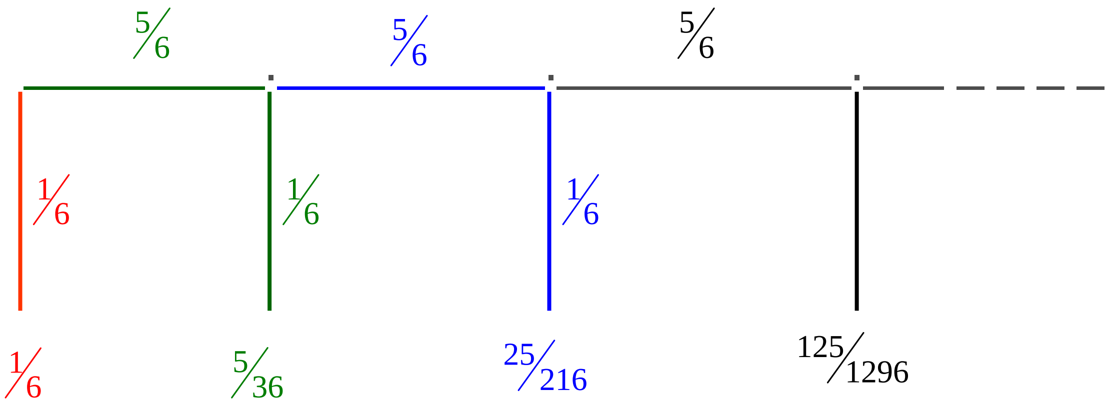
$$= (1-p)^{n-1}\,p$$

An easy example of a geometric distribution:

Probability of getting a 5 with a dice

$$P[N=1] = 1/6$$

$$P[N=2] = (1 - 1/6)\cdot 1/6$$



$$P[N=3] = \underbrace{(1 - 1/6)(1 - 1/6)}_{t-1}\cdot 1/6$$

$$= (1 - 1/6)^{3-1}\cdot 1/6$$

Exercise 6.2

A dike is designed to withstand a “1000-year return period flood”. How large is the probability that water will overflow the dike in the following conditions:

a) During a 10 year period, for the first time in a given year?
(for example, in the 10th year)

The probability that a flood occurs in the n^{th} year is

$$P[N=n] = \underbrace{(1-p)(1-p)\ldots(1-p)}_{n-1}\,p \quad \longleftarrow \text{Geometric distribution}$$
$$= (1-p)^{n-1}\,p$$

---

a) The event of overflow in the 10th year during a 10 year period may be described by a geometric distribution:

$$P(H_{overflow,1}) = (p)\cdot(1-p)^{n-1} = (0.001)\cdot(0.999)^{9} = 0.000991$$

Exercise 6.2

A dike is designed to withstand a “1000-year return period flood”. How large is the probability that water will overflow the dike in the following conditions:

b) During a 10 year period, twice?

Probability that in 10 years (*n* trials) you get 2 overflows (*y* successes)
→ Binomial distribution

$$P[Y = y] = \binom{n}{y} p^y (1 - p)^{n-y}$$

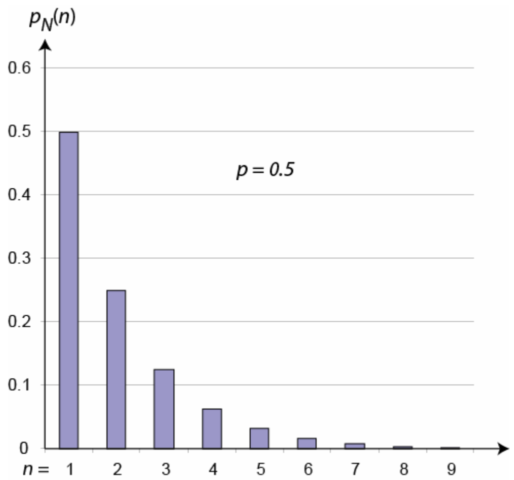
According to the Binomial distribution it is

$$P(H_{overflow,2}) = \frac{10!}{2! \cdot (10-2)!} (p)^2 \cdot (1-p)^{10-2}$$

Exercise 6.2

Review:

Geometric distribution
Time till first success

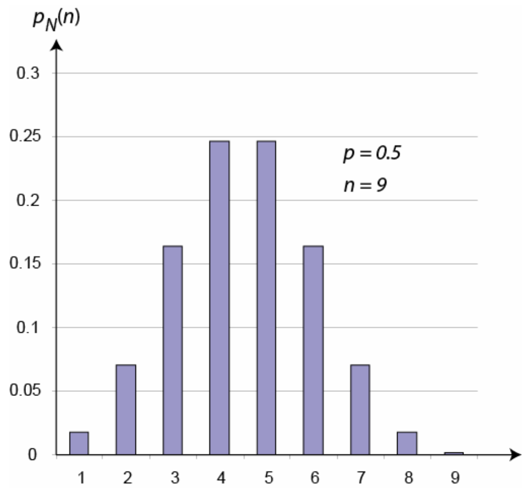


$$P[T = t] = (1-p)^{t-1}p$$

$$E[T] = \frac{1}{p}$$

$$Var[T] = \frac{1-p}{p^2}$$

Binomial distribution
Number of success



$$P[N = y] = \binom{n}{y}(1-p)^{n-y}p^y$$

$$E[N] = np$$

$$Var[N] = np(1-p)$$

A dike is designed to withstand a “1000-year return period flood”. How large is the probability that water will overflow the dike in the following conditions:

c) Will not overflow in a 10 year period?

Probability that in 10 years (*n* trials) you get 0 overflows (*y* successes)

→ Binomial distribution

$$P[Y = y] = \binom{n}{y} p^y (1-p)^{n-y}$$

$$P(H_{overflow,0}) = \frac{10!}{0! \cdot (10-0)!}(p)^0 \cdot (p-1)^{10-0} = (0.001)^0 \cdot (0.999)^{10} = \ldots$$

Exercise 6.2

A dike is designed to withstand a "1000-year return period flood". How large is the probability that water will overflow the dike in the following conditions:

d) During a 10 year period, at most once?

Probability that in 10 years (*n* trials) you get or 0 or 1 overflows (*y* successes)
→ Binomial distribution

$$P(H_{\max,1}) = P(H_{overflow,0}) + P(H_{overflow.,1})$$

$$P(H_{overflow,0}) = \frac{10!}{0!\cdot(10-0)!}(p)^0(p-1)^{10-0}$$

$$P(H_{overflow.,1}) = \frac{10!}{1!\cdot(10-1)!}(p)^1(p-1)^{10-1}$$

Exercise 6.2

A dike is designed to withstand a “1000-year return period flood”. How large is the probability that water will overflow the dike in the following conditions:

e) During a 100 year period, 10 times?

Probability that in 100 years (*n* trials) you get 10 overflows (*y* successes)
→ Binomial distribution

$$P(H_{overflow,10}) = \frac{100!}{10!(100-10)!}(p)^{10} \cdot (p-1)^{100-10}$$

Exercise 6.2

A dike is designed to withstand a “1000-year return period flood”. How large is the probability that water will overflow the dike in the following conditions:

g) During a 1000 year period, once or more often?

Probability that in 1000 years (n trials) you get any other result than 0 overflows (y successes)
→ Binomial distribution

$$P(H_{overflow,0}) = \frac{1000!}{0!(1000-0)!}(p)^0 \cdot (p-1)^{1000-0} = (0.001)^0 (0.999)^{1000} = 0.368$$

And the required probability is the probability of the complementary event:

$$P(H_{overflow,\geq 1}) = 1 - 0.368 = 0.632$$

Exercise 6.3 (Group exercise/or, during the exercise tutorial)

An environmental planning engineering company obtains a project in return for a project proposal with the success rate of 27%.

Assume that you have taken over this company and you need to make the business plan for the forthcoming years.

a. How large is the probability that the company will have at least one success after 12 project proposals?

b. How large is the probability that only the last of 10 project proposals is accepted?

c. How large is the probability that at most 2 out of 13 project proposals are accepted?

Exercise 6.3 (Group exercise/or, during the exercise tutorial)

What is given? Success rate = 27% → $p = 0.27$

a. How large is the probability that the company will have at least one success after 12 project proposals?

What is required: Time till first success – or number of successes?
How can you express “at least”?

b. How large is the probability that only the last of 10 project proposals is accepted?

What is required: Time till first success – or number of successes?

b. How large is the probability that at most 2 out of 13 project proposals are accepted?

What is required: Time till first success – or number of successes?
How can you express “at most”?