Midterm Review - Continuous RV

Continuous Random Variables

	Mean	Variance	PDF	CDF	MGF
Exponential	$\frac{1}{\beta}$	$\frac{1}{\beta^2}$	$\begin{cases} \beta\, e^{-\beta\, x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 - e^{-\beta\, x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\frac{\beta}{\beta - t}$
Gamma	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, e^{-\beta x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$	messy	$\left(1 - \frac{t}{\beta}\right)^{-\alpha}$
Chi Squared	γ	$2\,\gamma$	$\begin{cases} \frac{1}{\Gamma\left(\frac{\gamma}{2}\right) 2^{\gamma/2}} x^{\frac{\gamma}{2}-1}\, e^{-x/2} & x > 0 \\ 0 & \text{otherwise} \end{cases}$	messy	$(1 - 2\,t)^{-\gamma/2}$
Normal	μ	σ^2	$\frac{1}{\sqrt{2\,\pi}\;\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\frac{1}{2}\,\operatorname{erfc}\left(\frac{\mu - x}{\sqrt{2}\;\sigma}\right)$	$e^{\frac{\sigma^2\, t^2}{2} + \mu\, t}$
Standard Normal	0	1	$\frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{1}{2} x^2}$	$\frac{1}{2}\,\operatorname{erfc}\left(-\frac{x}{\sqrt{2}}\right)$	$e^{\frac{t^2}{2}}$
Weibull	μ_W	σ_W^2	$\begin{cases} \alpha\beta\, x^{\beta-1}\, e^{-\alpha\, x^\beta} & x > 0 \\ 0 & \text{otherwise} \end{cases}$	messy	messy
Uniform	$\frac{a+b}{2}$	$\frac{1}{12}\,(b-a)^2$	$\begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$	$\frac{e^{b\,t} - e^{a\,t}}{t\,(b-a)}$

- $\mu_W = \alpha^{-1/\beta}\,\Gamma\left(1 + \frac{1}{\beta}\right)$, $\sigma_W^2 = \alpha^{-2/\beta}\left(\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)^2\right)$.

Continuous Random Variable and its Properties

A ***continuous random variable*** is a map $X : S \to \mathbb{R}$ with a ***probability density function*** $f_X : \mathbb{R} \to \mathbb{R}$ such that

- $f_X(x) \geq 0$ for all $x \in \mathbb{R}$,
- $\int_{-\infty}^{\infty} f_X(x)\, d\,x = 1$.

And the ***cumulative distribution function*** is defined as

$$F_X(x) := P[X \leq\ x] = \int_{-\infty}^{x} f_X(y)\, d\,y\,,$$

which implies that $f_X(x) = F'_X(x)$.

The ***expectation*** has the similar properties,

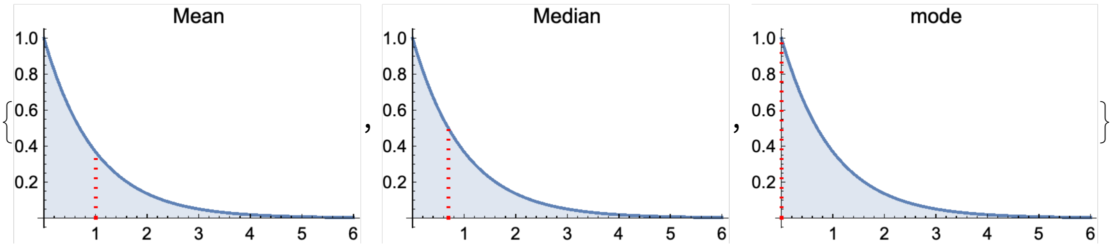
$$\mathrm{E}[X] := \int_{\mathbb{R}} x\, f_X(x)\, d\,x,\ \mathrm{E}[\varphi \circ X] = \int_{\mathbb{R}} \varphi(x)\, f_X(x)\, d\,x.$$

And the ***variance*** is given similarly, $\mathrm{Var}[X] = \mathrm{E}\left[\left(X - \mathrm{E}[X]\right)\right]^2 = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2$.

The ***median*** M_X is a number such that $P[X \leq M_X] = 0.5$.

The ***mode*** is a number at which f_X is maximum.

Out[∘]=



Exponential Distribution

Purpose: time until first arrival?

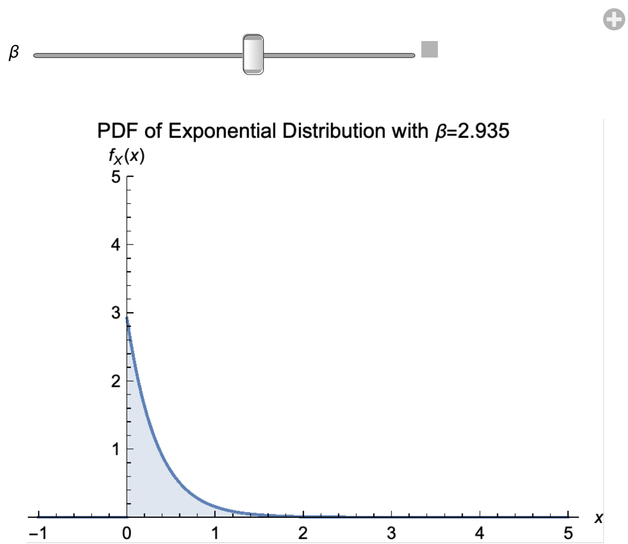
Parameter and properties:

- $\beta > 0$ is the rate of arrival.

Mean	Variance	PDF	CDF	MGF
$\frac{1}{\beta}$	$\frac{1}{\beta^2}$	$\begin{cases} \beta\, e^{-\beta x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 - e^{-\beta x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\frac{\beta}{\beta - t}$

- The exponential distribution is ***memoryless***, $P[X > x + s \,|\, X > x] = P[X > s]$.

Out[∘]=



Example:

Raindrops keep falling on my head at an average rate of 20 drops/min. What is the probability of having no raindrop falling on my head in a given 3 seconds time interval?

We can use $\beta = 1$ (drop/3 seconds). The probability of no drop (success) in the first 3 seconds is

$$P[X > 1] = 1 - \int_0^1 e^{-x}\, dx \; = 1 - \left(1 - \frac{1}{e}\right) = \frac{1}{e} = 36.7\,\%$$

Setting other values of β, such as $\beta = 20$ (drops/min) and calculating $P\left[X > \frac{1}{20}\right]$, will give the same result.

Given that no raindrop has fallen on my head in the past minute, what is the probability of having no raindrop on my head in another 3 seconds?

Still 36.7 % due to the memoryless property.

Gamma Distribution

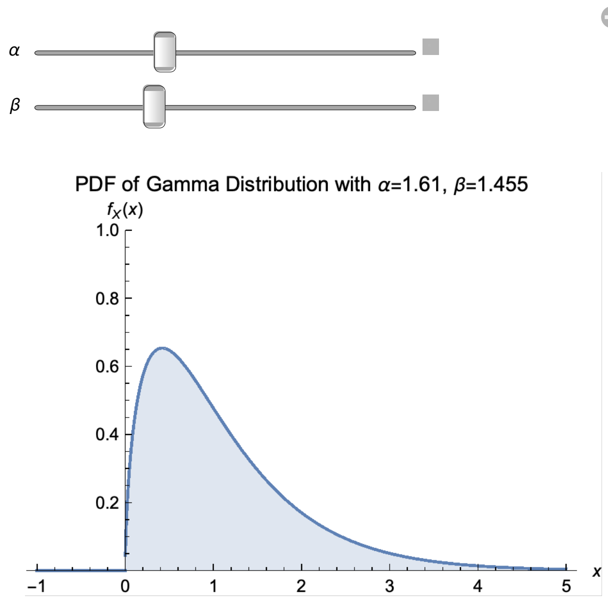
Purpose: time until α^{th} arrival?

Parameter and properties:

- $\alpha > 0$ is the number of arrivals you want.
- $\beta > 0$ is the rate of arrival.

Mean	Variance	PDF	MGF
$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$ where $\Gamma(\alpha)=\int_0^\infty z^{\alpha-1} e^{-z} \, dz$	$\left(1 - \frac{t}{\beta}\right)^{-\alpha}$

Out[•]=



Chi-squared Distribution

Purpose: how is the sum of squares of γ independent standard normal random variables distributed?

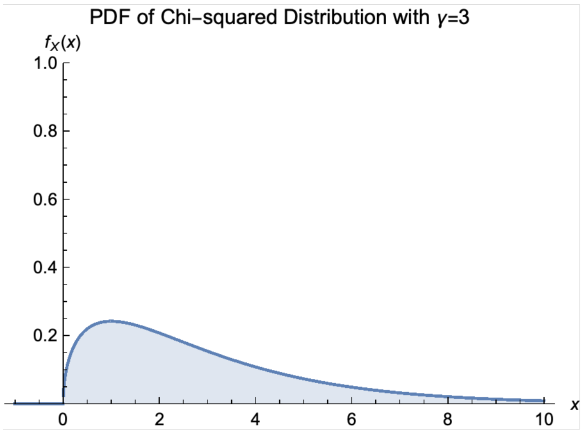
Parameter and properties:

- $\gamma \in \mathbb{N}^+$ is the ***degrees of freedom***.

Mean	Variance	PDF	MGF
γ	2γ	$\begin{cases} \frac{1}{\Gamma\left(\frac{\gamma}{2}\right) 2^{\gamma/2}} x^{\frac{\gamma}{2}-1} e^{-x/2} & x>0 \\ 0 & \text{True} \end{cases}$	$(1-2t)^{-\gamma/2}$

γ

Out[∘]=



Transformation of Random Variable

X is a continuous RV and $\varphi: \mathbb{R} \to \mathbb{R}$ is a strictly monotonic and differentiable function, then the PDF of $Y = \varphi \circ X$ is

$$f_Y(y) = \begin{cases} f_X\left(\varphi^{-1}(y)\right) \cdot \left| \frac{d\varphi^{-1}(y)}{dy} \right| & \text{for } y \in \operatorname{ran} \varphi \\ 0 & \text{otherwise} \end{cases}$$

Normal Distribution

Purpose: not sure which distribution? Then normal distribution!

A Useful Conclusion: $\int_{-\infty}^{\infty} e^{-x^2/c}\, dx = \sqrt{c\pi}$, where $c > 0$ is a constant.

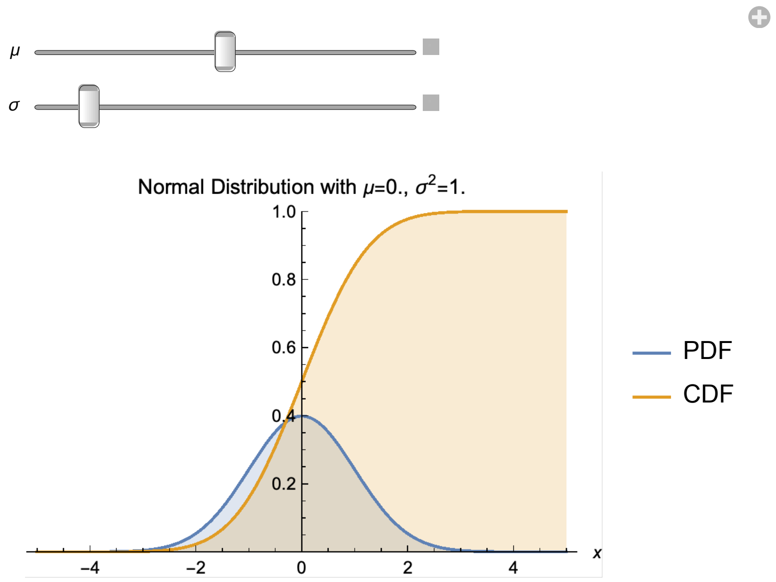
Parameter and properties:

- $\mu \in \mathbb{R}$ is the mean,
- $\sigma^2 > 0$ is the variance

Mean	Variance	PDF	CDF	MGF
μ	σ^2	$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\frac{1}{2}\operatorname{erfc}\left(\frac{\mu-x}{\sqrt{2}\,\sigma}\right)$ where $\operatorname{erfc}(z) := 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$	$e^{\frac{\sigma^2 t^2}{2}+\mu t}$

- Let X be normally distributed, then $Z := \frac{X-\mu}{\sigma}$ has ***standard normal distribution*** (normal distribution with $\mu = 0$, $\sigma = 1$).

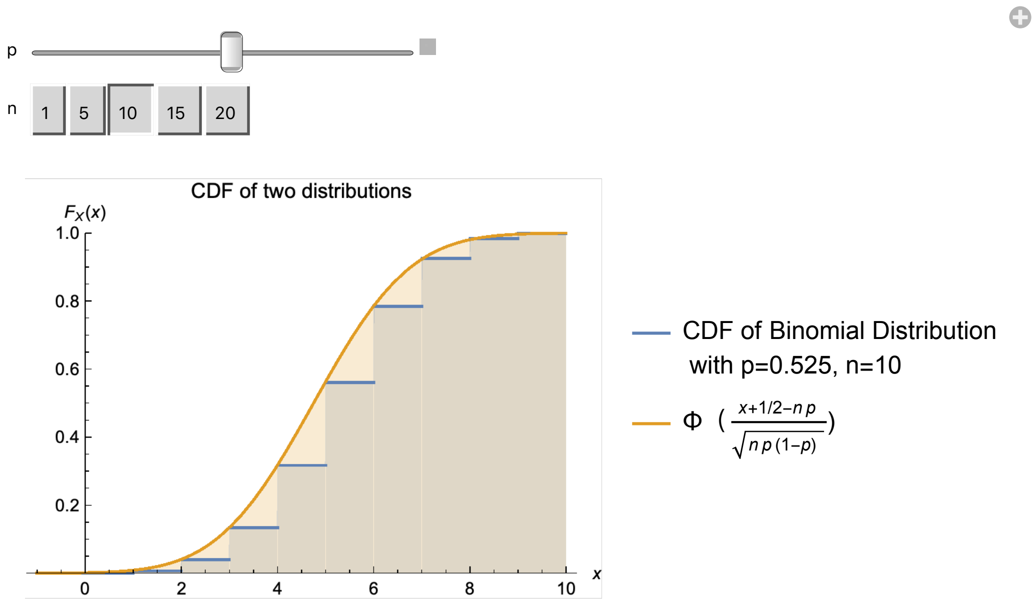
Out[∘]=



Approximation of Binomial CDF: if $\begin{cases} n > 10 & \text{if } p \text{ is close to } 1/2 \\ n\,p > 5 & \text{if } p \le 1/2 \\ n(1-p) > 5 & \text{if } p > 1/2 \end{cases}$, the CDF at $y \in \mathbb{N}$ can be approximated by normal distribution.

$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi\left(\frac{y + 1/2 - n\,p}{\sqrt{n\,p(1-p)}}\right)$$

Out[∘]=



Example: For a random variable X that follows a binomial distribution with $n = 100$, $p = 0.5$, use normal approximation to calculate $P[X \geq 40]$.

We have $P[X \geq 40] = 1 - P[X < 40] = 1 - P[X \leq 39] = 1 - \Phi\left(\frac{39 + 1/2 - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5^2}}\right)$

In[4]:= `1 - CDF[NormalDistribution[], (39 + 1 / 2 - 100 × 0.5) / Sqrt[100 × 0.5^2]]`

$$1-\mathrm{CDF}\left[\mathrm{NormalDistribution}[\,],\ \frac{39+1/2-100\times 0.5}{\sqrt{100\times 0.5^2}}\right]$$

Out[4]= 0.982136

Standard Normal Distribution

The standard normal distribution will enable you to find the value of CDF, $\Phi(x)$, simply by looking up the following table. There are many forms of standard normal table, and you can check them out here: https://en.wikipedia.org/wiki/Standard_normal_table. With the information we can calculate the following:

- $P[X \leq a] = P[X < a] = \Phi(a)$
- $P[a \leq X \leq b] = P[X \leq b] - P[X \leq a] = \Phi(b) - \Phi(a)$

Out[49]=

	0.	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

Estimation: From the table you can read (**how? which number in the table are you using?**)

$$P[-\sigma < X - \mu < \sigma] = 0.68$$
$$P[-2\,\sigma < X - \mu < 2\,\sigma] = 0.95$$
$$P[-3\,\sigma < X - \mu < 3\,\sigma] = 0.997$$

In[∘]:=
```
CDF[NormalDistribution[0, 1], 1] // N
```

Out[∘]=
```
0.841345
```

Example:

A machine in the soft drink factory produces bottles of drinks with mean $\mu = 500$ mL and standard deviation $\sigma = 1$ mL. How many mLs should a bottle have, so that 95% of the bottles produced by this machine have less drink than the bottle?

We need $Z = \frac{X-\mu}{\sigma} = \frac{X-500}{1} > 1.65 \Rightarrow X > 501.65$.

In[∘]:=
```
InverseCDF[NormalDistribution[500, 1], 0.95]
```

Out[∘]=
```
501.645
```

Chebyshev Inequality

Purpose: To (roughly) estimate the variance of a random variable.

Mathematical representation: for $k \in \mathbb{N}\backslash\{0\}$, $P[\,|X| \geq c] \leq \frac{\mathrm{E}[|X|^k]}{c^k}$.

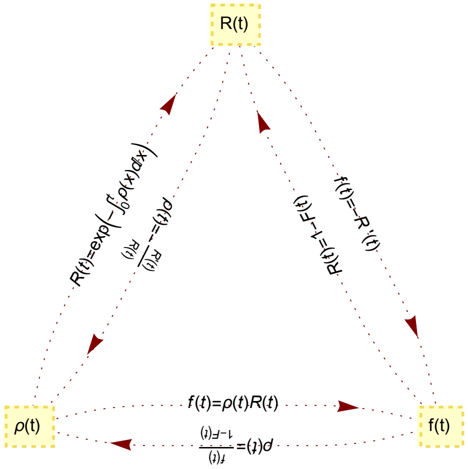
Application: $P[\,|X - \mu| \geq m\,\sigma] \leq \frac{1}{m^2}$.

Reliability

To study the reliability of a system, we have the following:

- ***Failure density*** $f(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t+\Delta t]}{\Delta t}$, the probability of failing at time t, (uniform, Weibull, ...)
- ***reliability function*** $R(t) = 1 - F(t)$, the probability of system still working at time t,
- ***hazard rate*** $\rho(t) = \lim_{\Delta t \to 0} \frac{P[t \leq T \leq t+\Delta t | T \geq t]}{\Delta t} = \frac{f(t)}{R(t)}$, the rate of failing at time t **given that it didn't fail before t**.

Out[•]=



- For series system, all components still working ⇒ system still working.
  $R_S(t) = P[\text{all components still working}] = \prod_{i=1}^{k} R_i(t)$.
- For parallel system, at least one component still working ⇒ system still working.
  $R_S(t) = P[\text{not all components broken}] = 1 - \prod_{i=1}^{k} (1 - R_i(t))$.

Example:

Your professor has announced that we will have a quiz on Monday class (just an example! not real!), but you don't know the exact time. What is the probability of not having a quiz in the first 45 minutes? Assume that a class is non-stop and 90 minutes long.

We can use uniform distribution to describe the "failure" density,

$$f(t) = \begin{cases} \frac{1}{90} & \text{if } 0 \le t \le 90 \\ 0 & \text{otherwise} \end{cases}$$

And the reliability at time $t = 45$ is

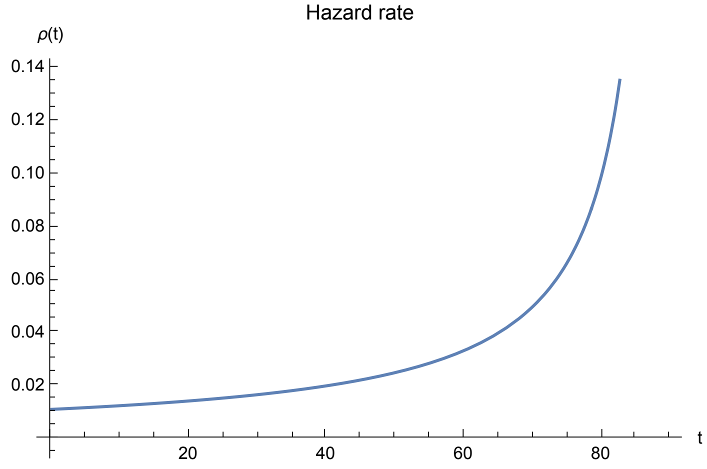
$$R(t) = 1 - F(t) = 1 - t / 90 = 0.5$$

what is the expected rate of your professor giving the quiz at the very next moment?

The hazard rate at $t = 45$ can be calculated as

$$\rho(t) = \frac{f(t)}{1-F(t)} = \frac{1/90}{1-t/90} = \frac{1/90}{1-45/90} = \frac{1}{45}$$

Out[ ]=



Weibull Distribution

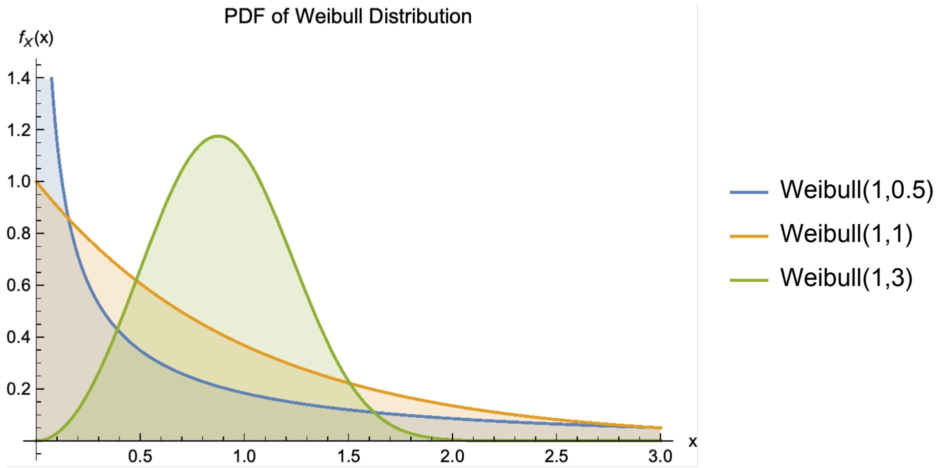
Purpose: To represent a failure density $f(t)$.

Parameters and properties:

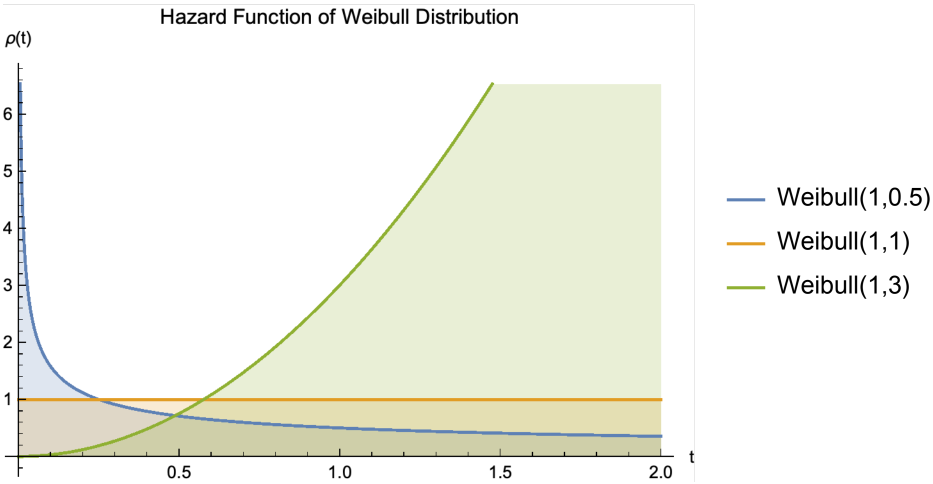
- $\alpha,\ \beta > 0$.
- the hazard rate is $\begin{cases} \text{constant} & \text{if } \beta = 1 \\ \text{increasing} & \text{if } \beta > 1 \\ \text{decreasing} & \text{if } \beta < 1 \end{cases}$.

Mean	Variance	PDF
$\alpha^{-1/\beta}\,\Gamma\left(1+\frac{1}{\beta}\right)$	$\alpha^{-2/\beta}\left(\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma\left(1+\frac{1}{\beta}\right)^2\right)$	$\begin{cases} \alpha\beta\, x^{\beta-1}\, e^{-\alpha\, x^{\beta}} & x>0 \\ 0 & \text{True} \end{cases}$

Out[ ]=



Out[•]=


Hazard Function of Weibull Distribution
ρ(t)
6
5
4
3
2
1
0.5
1.0
1.5
2.0
t
Weibull(1,0.5)
Weibull(1,1)
Weibull(1,3)