Lecture 7

David Armstrong

STATS 67 - UCI

Continuous Random Variables

- A continuous random variable can be an uncountable infinite possible values.
- Examples:

  X can be any number in the interval [0,1]. Thus $\mathbb{S}_X = [0, 1]$.

  Time it takes to drive to Las Vegas from UCI. Thus $\mathbb{S}_X = [3, \infty)$.

  Percent score on an Exam. Thus $\mathbb{S}_X = [0\%, 100\%]$.
- $f(x)$ is now called the *probability density function*.
- $f(x)$ does not represent $P(X = x)$ anymore.
- $P(X = x) = 0$ for all x in $\mathbb{S}_X$.
- The probability that a continuous random variable is equal to a single fixed number is 0.
- A continuous random variable takes on an uncountably infinite number of possible values.
- For a **discrete** random variable X that takes on a finite or countably infinite number of possible values, we determined that P(X = x) for all of the possible values of X, and called it the probability mass function (pmf).
- With a continuous random variable, we can only calculate probabilities of intervals such as $P(a < X < b)$.
- The pdf $f(x)$ is used to calculate the probability of intervals and moments.
- The pdf can be quantifying something that is proportional to the probability.

Continuous Random Variables

Let X be a continuous random variable with support $\mathbb{S}_X$ and probability density function $f(x)$.

- For $f(x)$ to be a valid pdf, the following must hold.
  - $f(x) \geq 0$ for all x in $\mathbb{S}_X$.
  - $\int\limits_{\mathbb{S}_X} f(x)dx = 1$.
- $E(X) = \int\limits_{\mathbb{S}_X} x\, f(x)dx$.
- $VAR(X) = \int\limits_{\mathbb{S}_X} (x - E(X))^2 f(x)dx$.
  - Note: We can still use the previous equation for the variance: $VAR(X) = E(X^2) - [E(X)]^2$.

Continuous Random Variables

Let X be a continuous random variable with support $\mathbb{S}_X$ and probability density function $f(x)$.

A few things to note.

- $P(X = x) = 0$.
- $P(X \leq x) = P(X < x)$.
  - $P(X \leq x) = P(X < x) + P(X = x) = P(X < x)$.
  - Example $P(X < 50) = P(X \leq 50)$, since P(X=50) is 0.
- Also, probability of intervals can be written using cdf's.

  $P(a < X < b) = F(b) - F(a)$.

The cumulative distribution function $F(x)$ is written as:

- $P(X < x) = P(X \leq x) = \int\limits_{l}^{x} f(u)du$.
  - Where l is the lower bound of the support of X, $\mathbb{S}_X$ (commonly it is $-\infty$).
- Note that $P(X < x) = F(X) = F(X) - F(l)$ where $F(l) = 0$.
- As a result, $\frac{d}{dx}F(x) = f(x)$.
- The derivative of the cumulative distribution function (cdf) is the probability distribution function (pdf).

Continuous Random Variables

Assume a distribution follows a bell shaped curve. Sketch each of the following situations.

- $P(X < b) = \int\limits_{x<b} f(x)dx$.

- $P(X > a) = \int\limits_{a<x} f(x)dx$.

- $P(a < X < b) = \int\limits_{a}^{b} f(x)dx$.

Normal Distribution

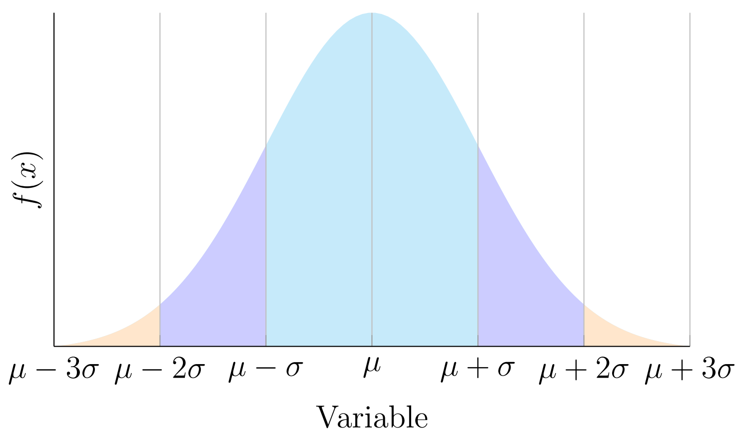
- Most common.
- Symmetric, unimodal, bell curve
- Gaussian distribution was named after Frederic Gauss, the first person to formalize its mathematical expression.
- $\mathbb{S}_X = (-\infty, \infty)$
- Say X follows a Normal distribution with parameters μ and σ.

  The location parameter is μ in $(-\infty, \infty)$

  The scale parameter is σ in $[0, \infty)$
- We write: $X \sim Normal(\mu, \sigma)$.
- Examples:
  - SAT scores
  - Heights of US adult males
  - The amount of time teenagers spend on the internet
  - Weights of babies

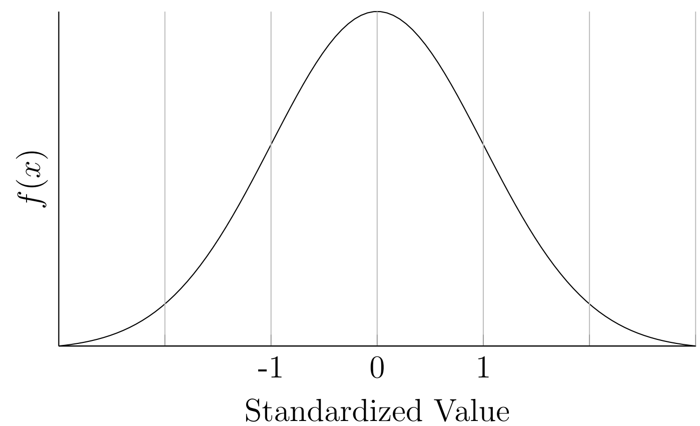
Normal Distribution

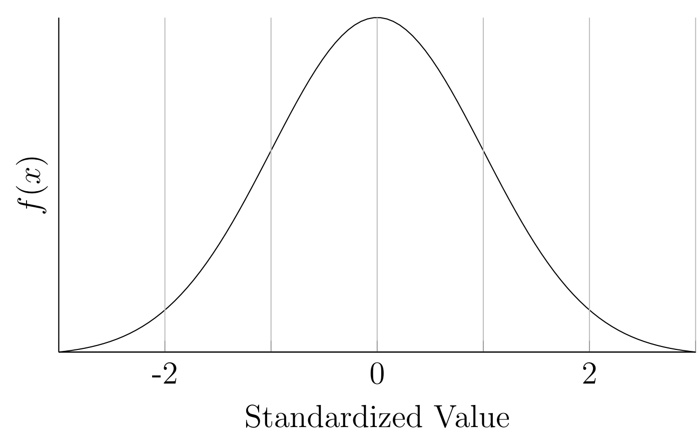
- Denoted $X \sim Normal(\mu, \sigma)$
- $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$.
  - We use this distribution for continuous variables that follow a bell shape curve.
- $\mathrm{F}(X) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$
- $\mathrm{E}(X)=\mu$
- $\mathrm{VAR}(X) = \sigma^2$
- Standardized score $Z = \dfrac{x-\mu}{\sigma}$

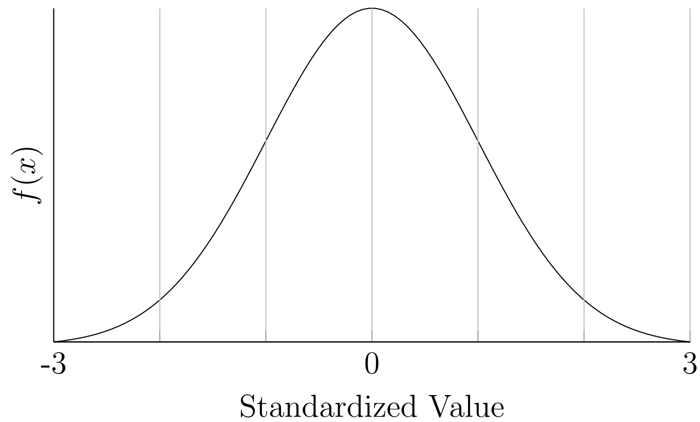


Empirical Rule

Here, we present a useful rule of thumb for the probability of falling within 1, 2, and 3 standard deviations of the mean in the normal distribution. This will be useful in a wide range of practical settings, especially when trying to make a quick estimate without R or Z-table. 68% will fall within one standard deviation of the mean, 95% within two standard deviations of the mean, and 99.7% within three standard deviations of the mean.







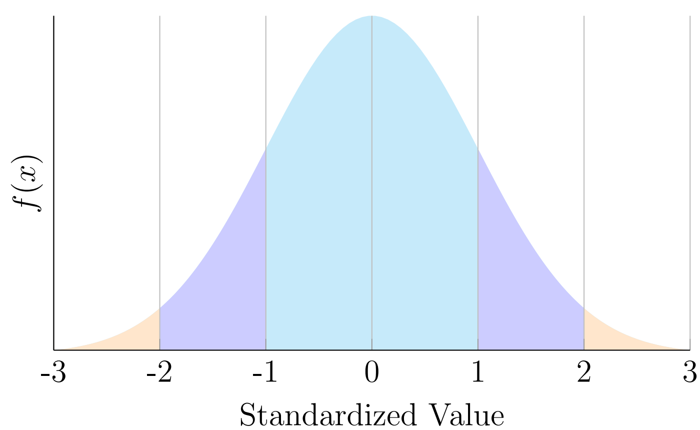
Example: Assume the poker winnings of an individual to be normally distributed with a mean of \$200 and a standard deviation of \$150.

- Use the empirical rule to determine the probability that the individual takes home more than \$350.

- 16% of the time, the individual will leave with how much in maximum winnings?

Standard Normal Distribution

- Denoted $Z \sim Normal(\mu = 0, \sigma = 1)$
- $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$.
  - We use this distribution to standardize the values of continuous variables that follow a bell shape curve.
- $\mathrm{F}(Z) = P(Z \leq z) = \int_{-\infty}^{z} f(z)dz$
- E(Z)=0
- VAR(Z) = 1



Probabilities of Z-scores $\leq$ -3.50 = .0001

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.40	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.30	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.20	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.10	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.00	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.90	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.80	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.70	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.60	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.50	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.40	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.30	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.20	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.10	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.50	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.40	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.20	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.00	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Probabilities of Z-scores $\geq 3.50 = .9999$

Example: What percent of a standard normal distribution $N(\mu = 0, \sigma = 1)$ is found in each region?

Be sure to draw a graph.

$P(Z < -1.35)$

$P(Z > -1.35)$

$P(-0.4 < Z < 1.5)$

Example: The distribution of SAT and ACT scores are both nearly normal.

	SAT	ACT
Mean	1500	21
SD	300	5

- Suppose Ann scored 1700 on her SAT and Tom scored 24 on his ACT. Who performed better?

- What is the probability someone scores above a 1550 on their SAT?

Example: Let X represent a random variable from $N(\mu = 3, \sigma = 2)$, and suppose we observe $x = 5.19$.

- Find the Z-score of x.

- Use the Z-score to determine how many standard deviations above or below the mean x falls.

R Code

To get the area to the left of a z-score:
$pnorm(z, mean = 0, sd = 1)$

To get the area to the right of a z-score:
$1 - pnorm(z, mean = 0, sd = 1)$

To get the area between two z-scores:
$pnorm(z_1, mean = 0, sd = 1) - pnorm(z_2, mean = 0, sd = 1)$

To get the z-score related to the lower tail (α):
$qnorm(\alpha, mean = 0, sd = 1)$

To get the z-score related to the upper tail:
$qnorm(1 - \alpha, mean = 0, sd = 1)$

Example: The length of time required to complete a college test is found to be normally distributed with mean 50 minutes and standard deviation 12 minutes.

a. When should the test be terminated if we wish to allow sufficient time for 90% of the students to complete the test?

b. What proportion of students will finish the test between 30 and 60 minutes?

c. What proportion of students will finish faster than 45 minutes?

Uniform Distribution

The *uniform distribution* is a distribution for a continuous random variable that can take on any value in an interval [a,b], with uniform density.

- Denoted $X \sim Uniform(a, b)$
- $\mathbb{S}_X = [a, b]$.
- $f(x) = \dfrac{1}{b-a}$.
  - We use this distribution for continuous variables where the values are all equally likely
- $\mathrm{F}(X) = P(X \leq x) = \int_a^x f(x)dx = \dfrac{x-a}{b-a}$
- $\mathrm{E}(X)=\dfrac{b+a}{2}$
- $\mathrm{VAR}(X) = \dfrac{(b-a)^2}{12}$
- The parameters of the distribution are a (lower bound) and b (upper bound).

This is a valid probability density function.

- For all x in [a,b], $f(x) \geq 0$.
- $\int\limits_{\mathbb{S}_X} f(x)dx = \int\limits_a^b \dfrac{1}{b-a}dx = \dfrac{x}{b-a}\Big|_{x=a}^{x=b} = \dfrac{b}{b-a} - \dfrac{a}{b-a} = 1$

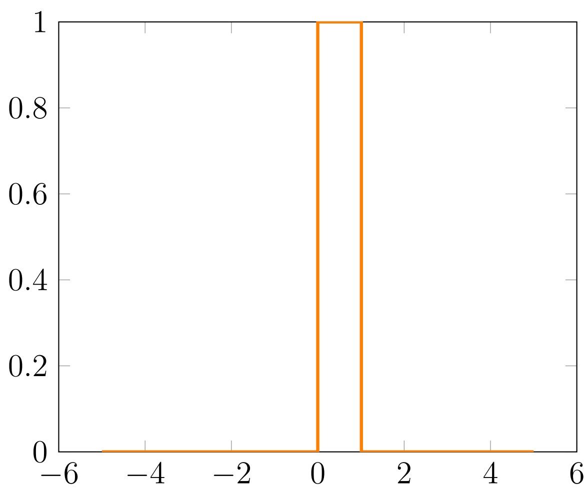
Uniform Distribution

Let X be a uniform continuous random variable on the interval $[a, b]$.

- $\mathbb{S}_X = [a, b]$.
- $f(x) = \dfrac{1}{b-a}$.

This is a valid probability density function.
Note: the CDF and pdf are only valid for x such that $a \leq x \leq b$

Ex: $X \sim Uniform(0, 1)$



Uniform Distribution

Let X be a uniform continuous random variable on the interval (a, b).

- $E(X) = \int\limits_{\mathbb{S}_X} x\, f(x)dx = \int\limits_a^b x\, \frac{1}{b-a} dx$
  $= \frac{x^2}{2(b-a)}\Bigg|_{x=a}^{x=b} = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}$

- The expected value of X is just the average of the two end points of the support.

- $E(X^2) = \int\limits_{\mathbb{S}_X} x^2\, f(x)dx = \int\limits_a^b x^2\, \frac{1}{b-a} dx$
  $= \frac{x^3}{3(b-a)}\Bigg|_{x=a}^{x=b} = \frac{b^3}{3(b-a)} - \frac{a^3}{3(b-a)} = \frac{(b^3-a^3)}{3(b-a)}$

Show that VAR(X)=$E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}$.

Uniform Distribution

Let X be a uniform continuous random variable on the interval (a,b).

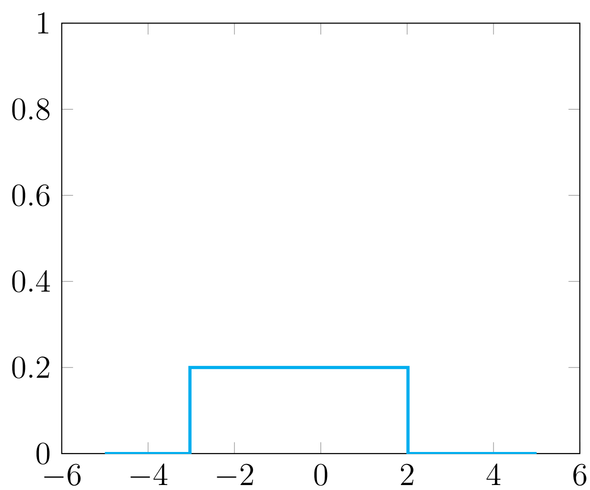
The cumulative distribution function is as follows:

- $F(x) = P(X < x) = \int\limits_{\tilde{x}<x} f(\tilde{x})d\tilde{x} = \int\limits_{a}^{x} \frac{1}{b-a}d\tilde{x}$
  $= \left.\frac{\tilde{x}}{(b-a)}\right|_{\tilde{x}=a}^{\tilde{x}=x} = \frac{x-a}{b-a}$

Thus for any x in (a,b), $F(x) = \frac{x-a}{b-a}$.

Note that if $x \leq a$ then F(x)=0 and if $x \geq b$ then F(x)=1.

Example: $X \sim Uniform(-3,2)$



Example Say you are waiting for the first car to cross an intersection. The waiting time is assumed to be any number between 0 and 13 minutes. The waiting time X is a uniform random variable on the interval [0,13].

- What is the probability that you wait less than 5 minutes?

- How much time do you expect to wait?

- Find the variance for the waiting time.

R Code

To get the area to the left of a Uniform(0,1) variable:
$punif(u, min = 0, max = 1)$

To get the area to the right of a Uniform(0,1) variable:
$1 - punif(u, min = 0, max = 1)$

To get the area between two values say c and d (c ¡ d):
$punif(d, min = 0, max = 1) - punif(c, min = 0, max = 1)$

To get the value of u related to the lower tail (α):
$qunif(\alpha, min = 0, max = 1)$

To get the value of u related to the upper tail:
$qunif(1 - \alpha, min = 0, max = 1)$

Exponential Distribution

Now we come to what is known as the *exponential distribution*.
It is used to model the the time between events in a Poisson process.

Let X be a random variable that follows an exponential distribution.

- $X \sim Exponential(\lambda)$
- The probability density function of X (pdf) is $f(x) = \lambda e^{-\lambda x}$.
- The expectation is $E(X) = \frac{1}{\lambda}$.
- The variance is $VAR(X) = \frac{1}{\lambda^2}$
- The cumulative distribution function is $F(x) = 1 - e^{-\lambda x}$.
- $P(a < X < b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}$
- λ is the parameter where $\lambda > 0$.
- The support of X is $\mathbb{S}_X = [0, \infty)$

Exponential Distribution

Let X be a random variable that follows an exponential distribution with parameter $\lambda > 0$.

$$\begin{aligned}
E(X) &= \int_0^\infty x f(x) dx \\
&= \int_0^\infty \lambda x e^{-\lambda x} dx \\
&= [-xe^{-\lambda x}]|_{x=0}^{x=\infty} + \int_0^\infty e^{-\lambda x} dx \\
&= (0-0) + \frac{1}{\lambda} x e^{-\lambda x}]|_{x=0}^{x=\infty} \\
&= 0 + \left(0 + \frac{1}{\lambda}\right) \\
&= \frac{1}{\lambda}
\end{aligned}$$

The parameter λ can be viewed as the expected time until the next event is observed.

Similarly we can show that $E(X^2) = \int_0^\infty \lambda x^2 e^{-\lambda x} dx = 2\left(\frac{1}{\lambda}\right)^2$

As a result, $VAR(X) = \frac{1}{\lambda^2}$.

Exponential Distribution

Let X be a random variable that follows an exponential distribution with parameter $\lambda > 0$.

The cumulative distribution function (cdf) F(x) is as follows.

$$
\begin{aligned}
F(x) &= \int_0^x f(u)du \\
&= \int_0^x -\lambda e^{-\lambda u} du \\
&= -e^{-\lambda u}|_{u=0}^{u=x} \\
&= -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}
\end{aligned}
$$

As a result: $P(a < X < b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}$.

Exponential Distribution

Let X follows an exponential distribution with a constant parameter $\lambda > 0$.

A property of this distribution is called its *memoryless* property.

- Notationally this is $P(X > x + t | X > x) = P(X > t)$ where $x, t > 0$.
- For example $P(X > 45 | X > 35) = P(X > 35 + 10 | X > 35) = P(X > 10)$

Given the wait time X for the next event is greater than x, the probability that the time X is greater than $x + t$ is just equal to the unconditional probability that X is greater than t.

Ex: The wait time (in minutes) to observe the next car that crosses an intersection follows an exponential distribution with $\lambda = \frac{1}{10}$.

a. What is the expected wait time for the next car to cross the intersection?

b. What is the variance of the wait time for the next car to cross the intersection?

c. Find the probability that the wait time is less than 5 minutes.

d. What is the probability that the wait time is between 4 and 6 minutes?

e. Now say we know (or that we are given, or condition on) that the wait time is more than 4 minutes. What is the probability that the wait time is more than 6 minutes?

R Code

To get the area to the left of a Exponential(1) variable:
$pexp(e, rate = 1)$

To get the area to the right of a Exponential(1) variable:
$1 - pexp(e, rate = 1)$

To get the area between two values say c and d (c ¡ d):
$pexp(d, rate = 1) - pexp(c, rate = 1)$

To get the value of e related to the lower tail (α):
$qexp(\alpha, rate = 1)$

To get the value of e related to the upper tail:
$qexp(1 - \alpha, rate = 1)$