STA13: Elementary Statistics

Lecture 10
Book Sections 4.4 - 4.5

Dmitriy Izyumin

February 5 2018

Discrete Random Variables

- Generic discrete distribution
  (values and probabilities in a table)
- Binomial
- others we won't cover
  - Poisson
  - Geometric
  - Hypergeometric
  - Negative Binomial
  - ... many others

Continuous Random Variables

Continuous random variables have (uncountably) infinitely many possible values.

Examples:

- Height, weight, blood pressure
- Laboratory measurement error
- Lifetime of an appliance component

Discrete vs. Continuous

Discrete Random Variables

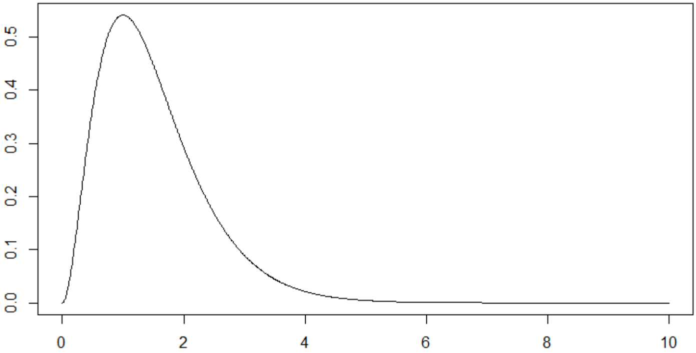
- Finitely or countably many values
- We can find $P(X = k)$ for any possible value k
- It is sometimes possible to put all the probabilities in a table (like in the generic case)
- It is sometimes possible to express all probabilities using a single formula (like for a Binomial r.v.)

Discrete vs. Continuous

Continuous Random Variables

- Uncountably many values
- The probability that X equals any k is 0:
  $P(X = k) = 0$ for any specific k
- We can find the probability that X is between a and b:
  $P(a \leq X \leq b)$
- Probabilities are summarized using a smooth function known as the probability density function (pdf)

Discrete vs. Continuous



We use a curve to model the distribution of a continuous r.v.

Probability Density Function

The pdf of a continuous r.v. X has the following properties.

- Denoted by $f_X(x)$, or simply $f(x)$
  (Note that the big X denotes the random variable, and the little x denotes the value plugged into the function)
- Does not have to be symmetric
- Can not have any negative values
  (curve is never below the x-axis)
- Total area under the curve is 1

Using the Density Function

Computing probabilities, mean and variance directly from the pdf requires calculus.

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

$$Var[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) dx$$

We will not be using calculus at all.

Some General Probability facts

Suppose X is a continuous random variable with pdf $f(x)$.

- $P(X = a) = 0$
- $P(X \leq a) = P(X < a)$

  (similarly for $\geq$ and $>$)
- $P(X \geq a) = 1 - P(X < a)$
- If $a \leq b$, then

  $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$

Continuous Random Variables

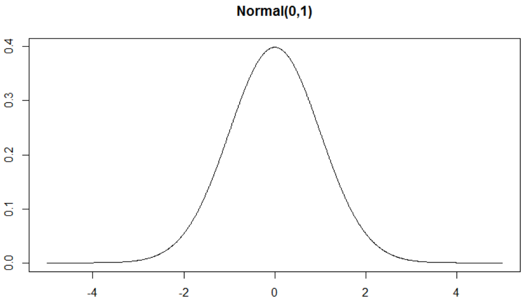
- There are many, many types of continuous random variables with different density functions.
- We will only cover a few of them.
- A very important continuous distribution is the normal distribution.

Normal Distribution

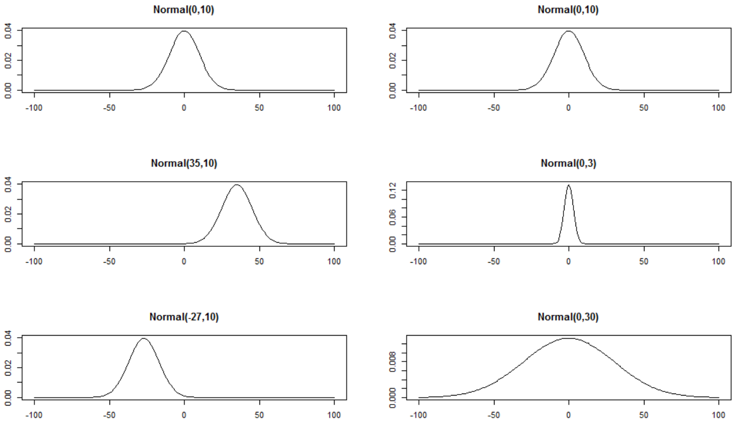
We say a random variable X is distributed Normal(μ, σ) if it has mean μ, standard deviation σ, and the pdf $f(x)$ below:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \text{ for } -\infty < x < \infty$$

The normal distribution has the recognizable bell shape:



Spread and Center



μ determines the center of the curve

σ determines the spread, or width

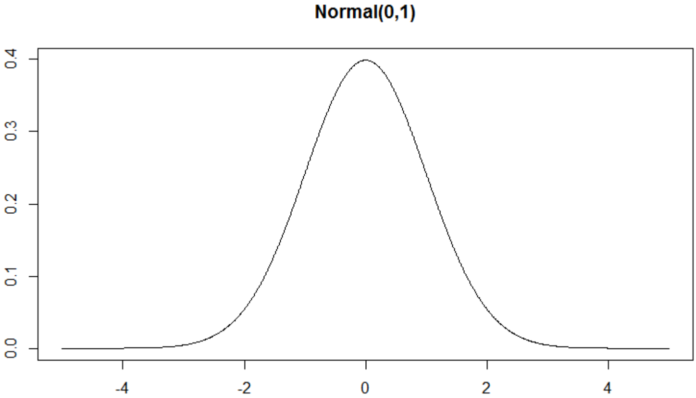
Mean and Variance

If X is distributed Normal(μ, σ), then

- μ is the mean of X
- σ is the standard deviation of X
- σ^2 is the variance of X
- We write $X \sim N(\mu, \sigma)$

Standard Normal

The $N(0,1)$ distribution is the standard normal distribution.



Standard Normal Table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295

- Has probabilites of the form $P(Z < a)$
- Can help us get probabilites like $P(Z > a)$ or $P(a \leq Z < b)$
- Full table is on Canvas

Using the Standard Normal Table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295

Example 1: Find $P(Z \leq 0.43)$.

- Find 0.4 and 0.03 in the margins
- The corresponding cell says 0.66640
- So $P(Z \leq 0.43) = 0.66640$

Using the Standard Normal Table

Some tricks:

- $P(Z < a) = P(Z \leq a)$ (because Z is continuous)
- $P(Z > a) = 1 - P(Z < a)$ (complement rule)
- $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$
- $P(Z < a) = P(Z > -a)$ (symmetry of $N(0, 1)$)

Using the Standard Normal Table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295

Example 2: Find $P(Z > 0.43)$.

- Rewrite $P(Z > 0.43) = 1 - P(Z < 0.43)$
- Get $P(Z < 0.43) = 0.66640$ as before
- So $P(Z > 0.43) = 1 - 0.66640 = 0.33360$

Using the Standard Normal Table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295

Example 3: Find $P(0.1 \leq Z \leq 0.43)$.

- $P(0.1 \leq Z \leq 0.43) = P(Z \leq 0.43) - P(Z \leq 0.1)$
- Get $P(Z \leq 0.43) = 0.66640$ as before
- Get $P(Z \leq 0.1) = 0.53983$ as before
- So $P(0.1 \leq Z \leq 0.43) = 0.66640 - 0.53983 = 0.12657$

Using the Standard Normal Table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295

Example 4: Find $P(Z > -0.43)$.

- Using symmetry, $P(Z > -0.43) = P(Z < 0.43)$
- Get $P(Z < 0.43) = 0.66640$ as before
- So $P(Z > -0.43)) = 0.66640$

Using the Standard Normal Table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295

Example 5: Find the 87th percentile of $N(0, 1)$

- Look in the table (not in margins!) for the number closest to 0.87.
- It is 0.87076
- Corresponding percentile (in margins) is 1.13.