Special Continuous Random Variables

Dason Kurkiewicz

Iowa State University

June 12, 2013

Outline

The normal (Gaussian) distribution

- A random variable X is Normal(μ, σ^2) if its pdf is:

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

- Using calculus, one can verify that:
  - $E(X) = \mu$
  - $\text{Var}(X) = \sigma^2$
- $\frac{X-\mu}{\sigma} \sim N(0,1)$, where N(0,1) is the *standard* normal distribution (mean 0, variance 1).

The standard normal distribution

- A standard normal random variable, usually called Z, has the pdf:

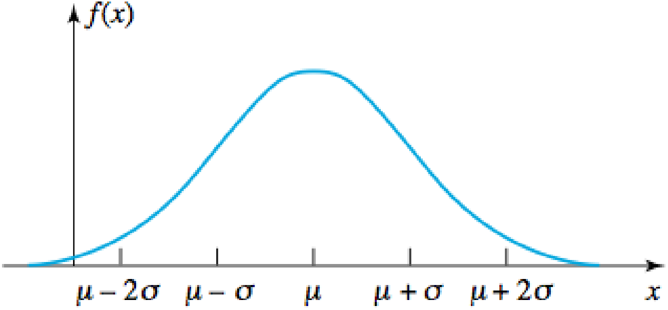
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- The standard normal pdf is usually denoted $\phi(z)$.
- The standard normal cdf is usually denoted $\Phi(z)$.

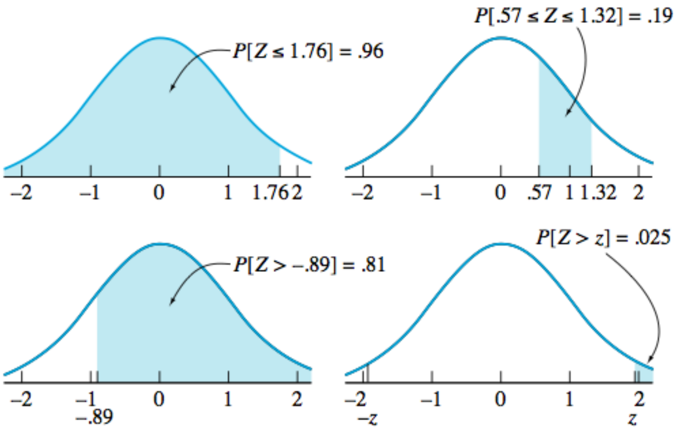
Uses of the normal distribution

- A normal random variable is (often) used to model measurement error, natural characteristics, and finite averages of many repeated, independent, identical trials.
- Examples:
  - Several kinds of measurement error.
  - The next blood pressure reading.
  - Corrosion resistance of carbon/carbon composites.
  - Mean width of the next 50 hexamine pellets.
  - Mean height of the next 30 students.
  - Your SAT score.
  - Total % yield of the next 40 runs of a chemical process.

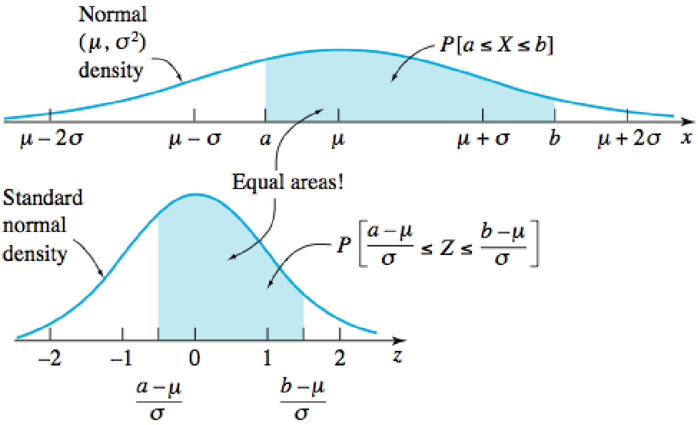
A look at the normal density: a bell curve



As usual, areas denote probabilities



The relationship between normal probabilities and standard normal probabilities.



Outline

Normal probabilities

- Since $Z = \frac{X-\mu}{\sigma}$ is *standard* normal probability values from X can be expressed as:

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$
$$= \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

- Unfortunately, the integral cannot be evaluated analytically. Instead, we use either:
  - A computer.
  - A standard normal probability table like the one in Table B.3 in Vardeman and Jobe.

Example: baby food

- J. Fisher, in his article ?Computer Assisted Net Weight Control? (Quality Progress, June 1983), discusses the filling of food containers with strained plums with tapioca by weight. The mean of the values portrayed is about 137.2 g, the standard deviation is about 1.6 g, and data look bell-shaped.
- Let W = the next fill weight. Then, $W \sim N(\mu = 137.2, \sigma^2 = (1.6)^2)$.
- Let's find the probability that the next jar contains less food by mass than it's supposed to (declared weight = 135.05 g).

$$\begin{aligned} P(W < 135.0) &= P\left(\frac{W - 137.2}{1.6} < \frac{135.05 - 137.2}{1.6}\right) \\ &= P(Z < -1.34) \\ &= \Phi(-1.34) \end{aligned}$$

- The approximate value of $\Phi(-1.34)$ is found to be 0.0901 in Table B.3.

The standard normal table

Standard Normal Cumulative Probabilities

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Your turn: using the standard normal table, calculate the following.

1. $P(X \leq 3)$, $X \sim N(2, 64)$
2. $P(X > 7)$, $X \sim N(6, 9)$
3. $P(|X - 1| > 0.5)$, $X \sim N(2, 4)$
4. $P(X$ is within 2 standard deviations of its mean.) $X \sim N(\mu, \sigma^2)$

Answers: normal probabilities

1. $P(X \leq 3)$, $X \sim N(2, 64)$

$$
\begin{aligned}
P(X \leq 3) &= P\left(Z \leq \frac{3-2}{\sqrt{64}} = 0.125\right) \\
&= \Phi(0.125) \\
&= 0.5478 \text{ from the standard normal table}
\end{aligned}
$$

Answers: normal probabilities

2. $P(X > 7)$, $X \sim N(6, 9)$

$$
\begin{aligned}
P(X > 7) &= P\left(Z > \frac{7-6}{\sqrt{9}} = 0.33\right) \\
&= 1 - P\left(Z \leq 0.33\right) \\
&= 1 - \Phi(0.33) \\
&= 1 - 0.6293 \text{ from the standard normal table} \\
&= 0.3707
\end{aligned}
$$

Answers: normal probabilities

3. $P(|X-1| > 0.5)$, $X \sim N(2,4)$

$$\begin{aligned}
P(|X-1| > 0.5) &= P(X-1 > 0.5 \text{ or } X-1 < -0.5) \\
&= P(X-1 > 0.5) + P(X-1 < -0.5) \\
&= P(X > 1.5) + P(X < 0.5) \\
&= P\left(\frac{X-2}{2} > \frac{1.5-2}{2}\right) + P\left(\frac{X-2}{2} < \frac{0.5-2}{2}\right) \\
&= P(Z > -0.25) + P(Z < -0.75) \\
&= 1 - P(Z \leq -0.25) + P(Z \leq -0.75) \\
&= 1 - \Phi(-0.25) + \Phi(-0.75) \\
&= 1 - 0.4013 + 0.2266 \text{ from the standard normal table} \\
&= 0.8253
\end{aligned}$$

Answers: normal probabilities

4. $P(X$ is within 2 standard deviations of its mean.) $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
P(|X-\mu| < 2\sigma) &= P(-2\sigma < X - \mu < 2\sigma) \\
&= P(\mu - 2\sigma < X < \mu + 2\sigma) \\
&= P\left(\frac{(\mu - 2\sigma) - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{(\mu + 2\sigma) - \mu}{\sigma}\right) \\
&= P(-2 < Z < 2) \\
&= P(Z < 2) - P(Z < -2) \\
&= \Phi(2) - \Phi(-2) \\
&= 0.9773 - 0.0228 \\
&= 0.9545
\end{aligned}
$$

Outline

Normal quantiles

- One can find standard normal quantiles by using the standard normal table in reverse.
- Example: for the jar weights $W \sim (137.2, 1.6^2)$ find $Q(0.1)$

$$
\begin{aligned}
0.1 &= P(X \leq Q(0.1)) \\
&= P\left(Z \leq \frac{Q(0.1) - 137.2}{1.6}\right) \\
&= \Phi\left(\frac{Q(0.1) - 137.2}{1.6}\right) \\
\Phi^{-1}(0.1) &= \frac{Q(0.1) - 137.2}{1.6} \\
Q(0.1) &= 137.2 + 1.6 \cdot \Phi^{-1}(0.1)
\end{aligned}
$$

$\Phi^{-1}(0.1) = -1.28$ from the standard normal table. Hence:

$$
\begin{aligned}
Q(0.1) &= 137.2 + 1.6(-1.28) \\
&= 135.152
\end{aligned}
$$

Finding $Q(0.1)$

Table B.3

Standard Normal Cumulative Probabilities

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Your turn: calculate the following:

1. $Q(0.95)$ of $X \sim N(9, 3)$
2. c such that $P(|X - 2| > c) = 0.01$, $X \sim N(2, 4)$
3. c such that $P(|X - \mu| < \sigma c) = 0.95$, $X \sim N(\mu, \sigma^2)$

Answers

1. $Q(0.95)$ for $X \sim N(9, 3)$

$$
\begin{aligned}
0.95 &= P(X \leq Q(0.95)) \\
&= P\left(\frac{X-9}{\sqrt{3}} < \frac{Q(0.95)-9}{\sqrt{3}}\right) \\
&= P\left(Z < \frac{Q(0.95)-9}{\sqrt{3}}\right) \\
0.95 &= \Phi\left(\frac{Q(0.95)-9}{\sqrt{3}}\right) \\
\Phi^{-1}(0.95) &= \frac{Q(0.95)-9}{\sqrt{3}} \\
Q(0.95) &= \sqrt{3} \cdot \Phi^{-1}(0.95) + 9 \\
&= \sqrt{3} \cdot (1.64) + 9 \quad \text{(from the std. normal table)} \\
&= 11.84
\end{aligned}
$$

Answers

2. c such that $P(|X-2|>c)=0.01$, $X\sim N(2.1,4)$

$$
\begin{aligned}
0.01 &= P(|X-2|>c) \\
&= P(X-2>c \text{ or } X-2<-c) \\
&= P(X-2>c)+P(X-2<-c) \\
&= P\left(\frac{X-2}{2}>\frac{c}{2}\right)+P\left(\frac{X-2}{2}<-\frac{c}{2}\right) \\
&= P\left(Z>\frac{c}{2}\right)+P\left(Z<-\frac{c}{2}\right) \\
&= P\left(Z<-\frac{c}{2}\right)+P\left(Z<-\frac{c}{2}\right) \quad (\phi(z) \text{ is symmetric about } 0) \\
&= 2P\left(Z<-\frac{c}{2}\right) \\
0.01 &= 2\Phi(-c/2) \\
0.005 &= \Phi(-c/2) \\
\Phi^{-1}(0.005) &= -c/2 \\
c &= -2\Phi^{-1}(0.005) \\
&= -2\cdot(-2.58) \quad (\text{using the standard normal table}) \\
&= 5.16
\end{aligned}
$$

Answers

3. c such that $P(|X-\mu| < \sigma c) = 0.95$, $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
0.95 &= P(|X-\mu| < \sigma c) \\
&= P(-\sigma c < X - \mu < \sigma c) \\
&= P\left(-c < \frac{X-\mu}{\sigma} < c\right) \\
&= P(-c < Z < c) \\
&= P(Z < c) - P(Z < -c) \\
&= (1 - P(Z > c)) - P(Z < -c) \\
&= (1 - P(Z < -c)) - P(Z < -c) \\
&\quad \text{(since } \phi(z) \text{ is symmetric about 0)} \\
&= 1 - 2P(Z < -c) \\
0.95 &= 1 - 2\Phi(-c) \\
0.05 &= 2\Phi(-c) \\
c &= -\Phi^{-1}(0.025) \\
&= -(-1.96) \quad \text{from the standard normal table} \\
&= 1.96
\end{aligned}
$$

Outline

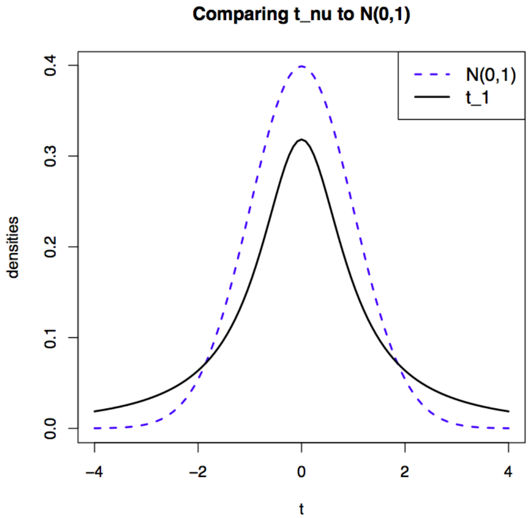
The Student t distribution

- A random variable T has a t_ν distribution – that is, a t distribution with ν **degrees of freedom** – if its pdf is:

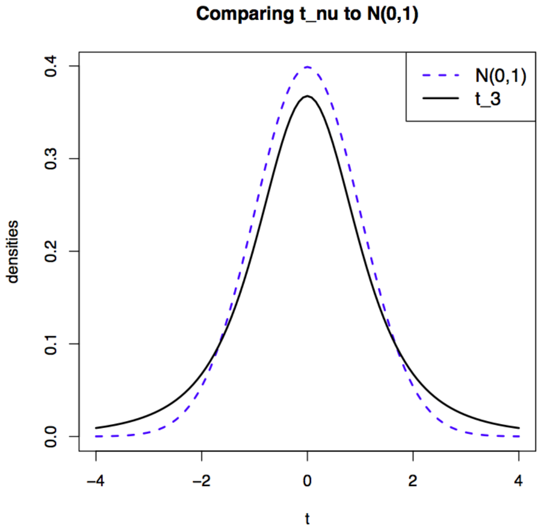
$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\nu\pi)^{\frac{1}{2}}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty$$

- $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$
- We use the t table (Table B.4 in Vardeman and Jobe) to calculate quantiles and probabilities.
- Like the standard normal distribution, the t distribution is mound-shaped and symmetric about 0.
- The t distribution has fatter tails than the normal, but approaches the shape of the normal as $\nu \to \infty$

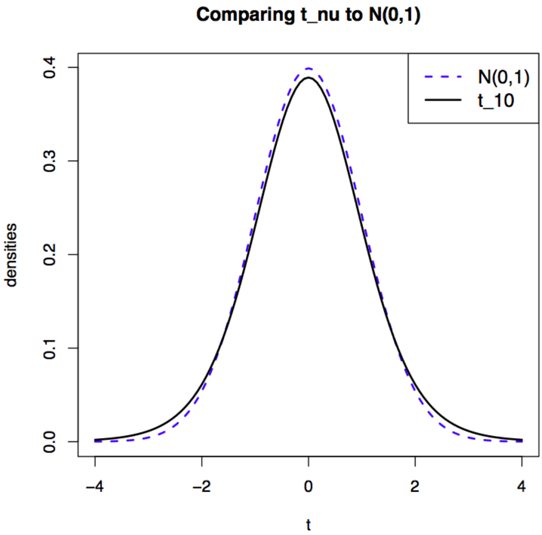
A look at the t_ν density



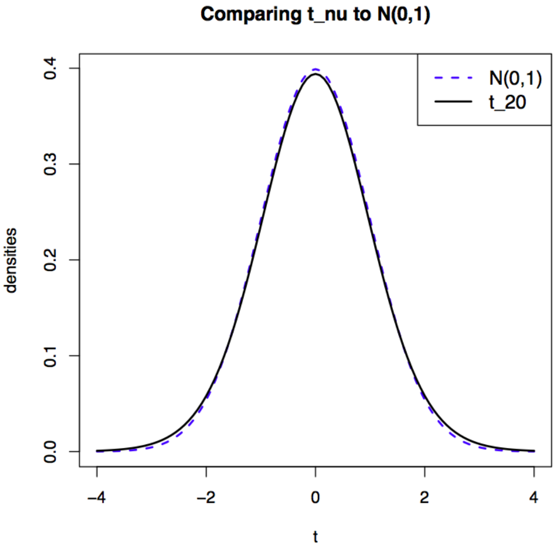
A look at the t_ν density



A look at the t_ν density



A look at the t_ν density



Find probabilities and quantiles of t_ν with the t table.

- Say $T \sim t_5$. $P(T \leq 1.476) = 0.9$

Table B.4
t Distribution Quantiles

ν	$Q(.9)$	$Q(.95)$	$Q(.975)$	$Q(.99)$	$Q(.995)$	$Q(.999)$	$Q(.9995)$
1	3.078	6.314	12.706	31.821	63.657	318.317	636.607
2	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869

- You can find quantiles labeled in the top row.

Outline

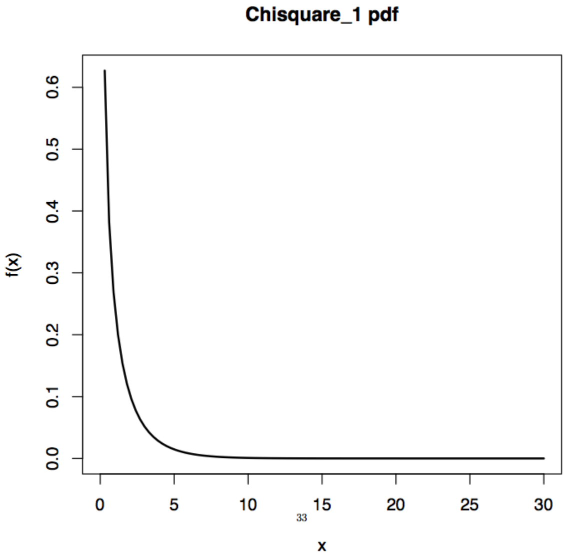
The chi-square distribution

- A random variable $S \sim \chi^2_\nu$ (is chi-square with ν **degrees of freedom**) if its pdf is:

$$f(x) = \begin{cases} 0 & : x \leq 0 \\ \frac{1}{\Gamma(\nu/2)2^{\nu/2}} \cdot x^{\nu/2-1} \cdot e^{-x/2} & : 0 < x < \infty \end{cases}$$

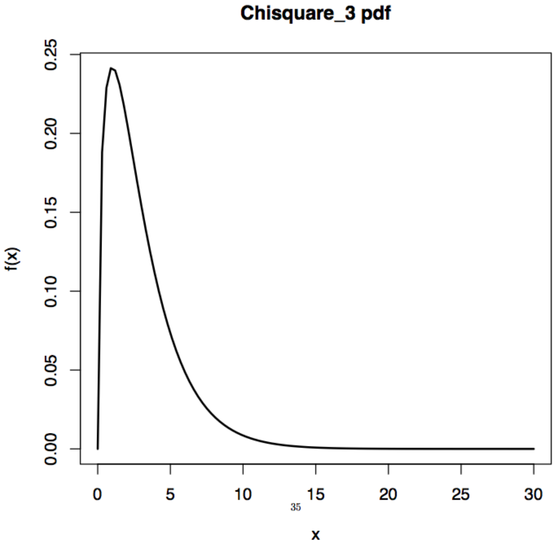
- Use Table B.5 in Vardeman and Jobe to find chisquare probabilities and quantiles.
- If $Z \sim N(0,1)$ then $Z^2 \sim \chi^2_1$
- The sum of ν independent χ^2_1 random variables has a $S \sim \chi^2_\nu$ distribution.
- The chi-square distribution is not symmetric (but gets closer to bell-shaped for large values of ν).

A look at the chi-square density



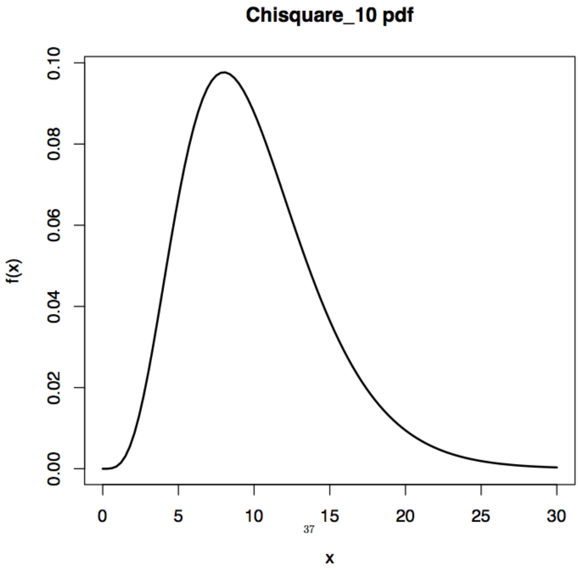
33

A look at the chi-square denstiy



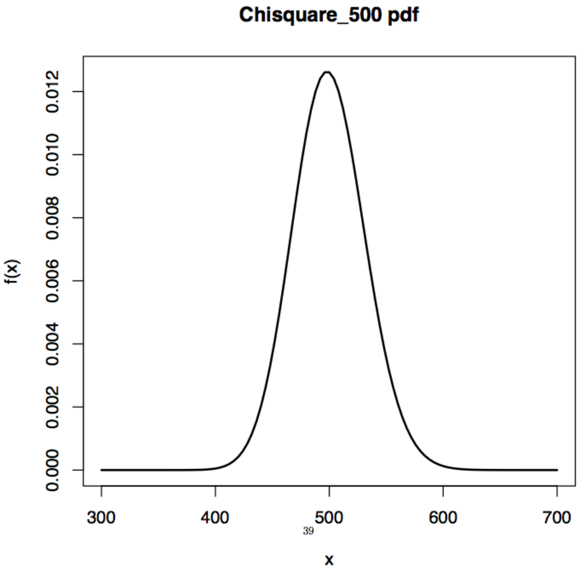
35

A look at the chi-square density



37

A look at the chi-square density



39

Use Table B.5 to find chi-square probabilities and quantiles.

- Q(0.9) of χ^2_6 is 10.645.

Table B.5
Chi-Square Distribution Quantiles

ν	$Q(.005)$	$Q(.01)$	$Q(.025)$	$Q(.05)$	$Q(.1)$	$Q(.9)$	$Q(.95)$	$Q(.975)$	$Q(.99)$	$Q(.995)$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548

Outline

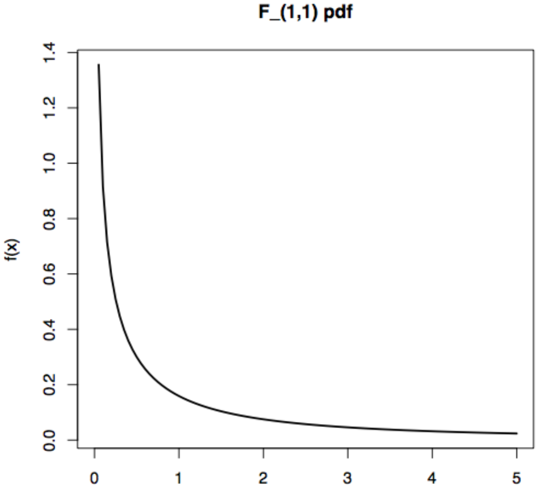
The F distribution

- X has an F_{ν_1,ν_2} distribution if it has pdf:

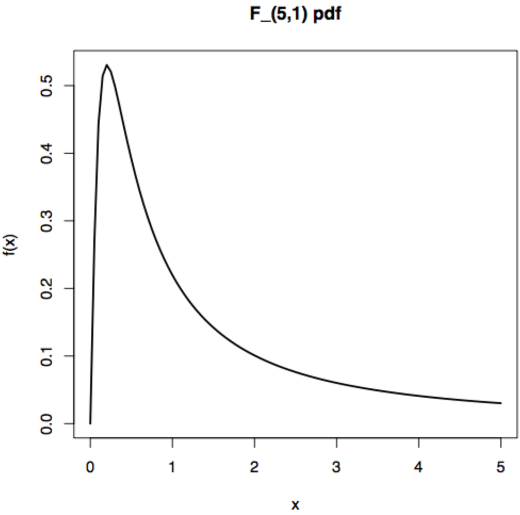
$$f(x) = \begin{cases} 0 & : x \le 0 \\ \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \dfrac{x^{\nu_1/2-1}}{[1+(\nu_1/\nu_2)x]^{(\nu_1+\nu_2)/2}} & : 0 < x < \infty \end{cases}$$

- An F_{ν_1,ν_2} random variable is a $\chi^2_{\nu_1}$ RV divided by an independent $\chi^2_{\nu_2}$ RV. That's why ν_1 is the **numerator degrees of freedom** and ν_2 is the **denominator degrees of freedom**.
- Use Tables B.6A-B.6E to find probabilities and quantiles.

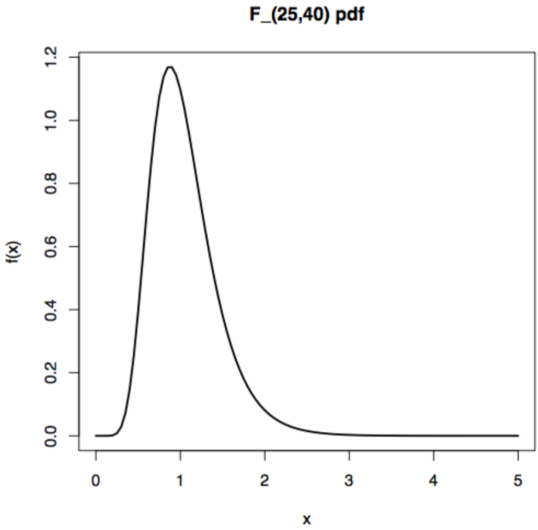
A look at the F density



A look at the F density



A look at the F density



Find probabilities and quantiles of the F distribution with Tables B.6A-B.6E

- The 0.99 quantile of the $F_{4,5}$ distribution is 11.39.

Table B.6D
F Distribution .99 Quantiles

ν_2 (Denominator Degrees of Freedom)	ν_1 (Numerator Degrees of Freedom) 1	2	3	4	5	6	7	8	9	10
1	4052	4999	5403	5625	5764	5859	5929	5981	6023	6055
2	98.51	99.00	99.17	99.25	99.30	99.33	99.35	99.38	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05

Outline

Special notation of quantiles

1. $Q(p)$ for $N(0, 1)$ is often denoted z_p.
2. $Q(p)$ for t_ν is often denoted $t_{\nu,p}$.
3. $Q(p)$ for χ^2_ν is often denoted $\chi^2_{\nu,p}$.
4. $Q(p)$ for F_{ν_1,ν_2} is often denoted $F_{\nu_1,\nu_2,p}$.