HW 4.3

\[ X, Y \] are independent random variables. \( \Rightarrow f(x,y) = f_X(x) \cdot f_Y(y) \)

Gaussian

\[ \bar{X} = 1 \quad \bar{Y} = 2 \]

\[ \sigma_X^2 = 1 \quad \sigma_Y^2 = 4 \]

Recall: Gaussian only.

\[ F(x) = \phi \left( \frac{x-\bar{X}}{\sigma_X} \right) \]

\[ \phi(-y) = 1 - \phi(y) \]

\[ P_X(X < x) = F(x) \]

\[ \text{Density Function} \]

\[ P_Y(Y > y) = 1 - P_Y(Y < y) \]

\[ P_X(X < x)P_Y(Y < y) = P_X(X < x) + P_Y(Y > y) - P_X(X < x)P_Y(Y < y) \]

\[ = P_X(X < x)P_Y(Y < y) + P_X(X > x)P_Y(Y > y) \]

\[ = F_X(x)F_Y(y) + [1 - F_X(x)][1 - F_Y(y)] \]

\[ = \frac{\phi(0-1)}{\phi(-1)} \frac{\phi(0-2)}{\phi(-1)} + \frac{[1 - \phi(-1)][1 - \phi(1)]}{\phi(-1)\phi(1)} \]

\[ = [1 - \phi(1)]^2 + [\phi(y)]^2 \]

\[ = [1 - 0.8413]^2 + (0.8413)^2 = 0.7329 \]

\[ \approx 0.733 \]

App D pg 432

4.4

Data

\[ \begin{align*}
X &: \bar{X} = 0 \quad \sigma_X^2 = 1 \\
Y &: \bar{Y} = 0 \quad \sigma_Y^2 = 1 \\
Z &: \bar{Z} = 0 \quad \sigma_Z^2 = 1
\end{align*} \]

\[ X \perp Y \quad \text{uncorrelated} \quad \Rightarrow P_{XY} = 0 \]

\[ X \perp Z \quad \Rightarrow P_{XZ} = \frac{1}{2} \]

\[ Y \perp Z \quad \Rightarrow P_{YZ} = \frac{1}{2} \]

a) \( \sigma_W^2 = ? \)

\[ W = X + Y + Z \quad \Rightarrow \bar{W} = \bar{X} + \bar{Y} + \bar{Z} = 0 \]

\[ \sigma_W^2 = \frac{W^2}{\bar{W}^2} = E(W^2) = E[X^2 + Y^2 + Z^2 + 2XY + 2YZ + 2XZ] \]


Recall

\[ E \left[ \frac{X - \bar{X}}{\sigma_X} \cdot \frac{Y - \bar{Y}}{\sigma_Y} \right] = E \left[ \frac{XY}{\sigma_X \sigma_Y} \right] \]

\[ \Rightarrow \sigma_X^2 = \bar{X}^2 - \bar{X}^2 = \bar{X}^2 = E[X^2] \]

\[ \sigma_Y^2 = \bar{Y}^2 - \bar{Y}^2 = \bar{Y}^2 = E[Y^2] \]

\[ \sigma_Z^2 = \bar{Z}^2 - \bar{Z}^2 = \bar{Z}^2 = E[Z^2] \]

\[ \sigma_{XY} = \bar{X} \cdot \bar{Y} = \bar{X} \cdot \bar{Y} \]

\[ \sigma_{XZ} = \bar{X} \cdot \bar{Z} = \bar{X} \cdot \bar{Z} \]

\[ \sigma_{YZ} = \bar{Y} \cdot \bar{Z} = \bar{Y} \cdot \bar{Z} \]

\[ \Rightarrow \sigma_W^2 = \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 + 2\sigma_{XY} + 2\sigma_{YZ} + 2\sigma_{XZ} \]

[\text{Plugging in values}]

\[ \sigma_W^2 = 3 \]
b) \[ E[XW] = E[X^2 + XY + XZ] = \frac{\sigma_x^2}{\sigma_w} + \frac{1}{1} \frac{\rho_{yx} \sigma_y \sigma_z}{\sigma_w} + \frac{1}{2} \frac{\rho_{xz} \sigma_x \sigma_z}{\sigma_w} = \frac{3}{2} \]

Since \[ \rho_{xw} = E \left[ \frac{X - \bar{X}}{\sigma_x} \frac{W - \bar{W}}{\sigma_w} \right] = E \left[ \frac{XW}{\sigma_x \sigma_w} \right] \]

\[ \rho_{xw} = \frac{3/2}{\sigma_x \sigma_w} = \frac{3/2}{1 \cdot \sqrt{2}} = \frac{\sqrt{3}}{2} \]

\[ \rho_{y+2} = ? \]

\[ E[W(Y+2)] = E[(X+Y+Z)(Y+Z)] = E[XY + XZ + Y^2 + 2YZ + Z^2] \]

\[ = \frac{\rho_{yx} \sigma_y \sigma_x + \rho_{xz} \sigma_z \sigma_x + \sigma_x^2}{0} + \frac{2 \rho_{yz} \sigma_z \sigma_x}{\frac{2}{1}} \]

\[ \sigma_{y+2} = \frac{2}{2} - \frac{1}{2} \cdot 1.1 = \frac{3}{2} \]

\[ \sigma_{y+2} = \frac{(Y+Z)^2 - \bar{Y}^2}{\frac{2}{1}} \]

\[ = \frac{Y^2 + Z^2 + 2YZ}{2} = \frac{Y^2 + Z^2 + 2 \rho_{yz} \sigma_y \sigma_z}{\sigma_y} \]

\[ \frac{E[Y^2]}{\sigma_y} = \frac{2}{\sigma_y} \]

\[ \frac{\rho_{w(y+2)}}{\sigma_{y+2}} = \frac{3/2}{\sqrt{3} \cdot 1.1} = \frac{\sqrt{3}}{2} \]
Ch 4 Sampling Theory & Random Variables (cont.)

Last time:  Mean of sample mean \( \bar{X} = \bar{X} \)

\( \bar{X} \) = mean of population

\( \bar{X} \) = mean of population

Variance of sample mean: \( \text{Var} (\bar{X}) = \frac{\sigma_x^2}{n} \)

\( \sigma_x^2 \) = variance of population

\( n \) = size of sample

\( N \) = size of population

\( n \) = size of sample

\( N \) = size of population

\( \Rightarrow \text{Mean of sample variance} = \bar{S}^2 = \frac{n-1}{n} \sigma_x^2 \)

\( \bar{S}^2 \) = mean of sample variance

\( \sigma_x^2 \) = variance of population

If we define the "unbiased" sample variance:

\( \bar{S}^2 = \sigma_x^2 \)

Then:

\( \bar{S}^2 = \frac{n-1}{n} \bar{S}^2 \)

\( \bar{S}^2 \) = mean of unbiased sample variance

\( \bar{S}^2 \) = mean of unbiased sample variance

4.4: Sampling Distribution & Confidence Interval:

Example 4.4.2:

A very large population of resistors with \( \bar{R} = 100 \Omega \) (mean value of population) and \( S = 4 \Omega \) (sample variance). What are the interval limits on the sample mean for a confidence level of 95%?

A distinction:

Sample mean: \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \)

value of one random variable

A random variable

\[ n > 30 \to \bar{X} \text{ follows a Gaussian distribution} \]

\[ \bar{X} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]

\[ n < 30 \to x_i \text{ follows a Student's } T \text{ distribution} \]

\[ T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]

\[ \text{degree of freedom} = n - 1 \]

HW5 Ch 4: 2.4, 2.7, 3.3, 3.4 (due 4/17)
\[ n > 30 \rightarrow X_i \text{ follows a Gaussian distribution} \]

\[ z = \frac{\bar{X} - \mu}{\sigma_x / \sqrt{n}} \quad (\text{Recall: } \text{var}(\bar{X}) = \frac{\sigma_x^2}{n} \rightarrow \text{st. dev. of sample mean is } \frac{\sigma_x}{\sqrt{n}}) \]

\[ \overline{z} = \frac{\overline{X} - \bar{X}}{\frac{\sigma_x}{\sqrt{n}}} = 0 \quad (\text{Recall: } \overline{X} = \bar{X}) \]

\[ \text{var}(\bar{X}) = \overline{X^2} - \bar{X}^2 \]

\[ \text{var}(z) = \overline{z^2} - \overline{z}^2 = \frac{1}{n} \left( \frac{\overline{X}^2}{\sigma_x^2} - 2 \frac{\overline{X} \bar{X}}{\sigma_x^2} + \bar{X}^2 \right) \]

\[ = \frac{\left( \frac{\overline{X}}{\sigma_x} \right)^2}{\frac{\sigma_x^2}{n}} - \frac{2 \overline{X} \bar{X}}{\sigma_x^2} + \frac{\bar{X}^2}{\frac{\sigma_x^2}{n}} = \frac{\text{var}(\bar{X})}{\left( \frac{\sigma_x}{\sqrt{n}} \right)^2} = \frac{\sigma_x^2}{n} \]

Recall: \[ \text{var}(\bar{X}) = \frac{\sigma_x^2}{n} \text{ (large population)} \]

\[ z = \frac{\bar{X} - \mu}{\frac{\sigma_x}{\sqrt{n}}} \text{ has zero mean and unit variance, follows a Gaussian distribution. Good when } n > 30 \]
Why interval limits and confidence level? What is the connection?

Connected through a probability distribution:

Interval \((-x_2, x_2)\) is connected with a probability described by the area under the curve (shaded area)

Higher probability → higher confidence level.

Each probability or confidence level is connected with certain interval limits, and vice versa. For example, for a confidence level of 95%, what are the limits of the interval for \(z\)?

With this area under the Gaussian density function for \(z\) \((\bar{z} = 0; \sigma^2 = 1)\), what are the limits \(\pm k\) for the interval in \(z\) \((-k, k)\)?

Gaussian density function with \(\bar{z} = 0; \sigma^2 = 1\)

Two-sided test:

\[
P_z(-k \leq z \leq k) = 0.95 \implies \frac{F(k) - F(-k)}{F(z)} = \frac{\Phi(k) - \Phi(-k)}{0.95}
\]

Recall: Gaussian: \(F(z) = \Phi(\frac{z - \bar{z}}{\sigma}) = 1 - Q(\frac{z - \bar{z}}{\sigma})\)

\(\Phi(\cdot) = 1 - \Phi(\cdot)\)

\(\Phi(k) = \frac{1}{2} \approx 0.975\)
Look in App D: "Table for Normal Prob. Dist. Function \( \Phi \)." Find what \( k \) if \( \Phi(k) = 0.975 \). \( k = 1.96 \). 

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Limits for interval in ( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>( \pm 1.96 )</td>
</tr>
<tr>
<td>90%</td>
<td>( \pm 1.645 )</td>
</tr>
<tr>
<td>99%</td>
<td>( \pm 2.58 )</td>
</tr>
<tr>
<td>99.9%</td>
<td>( \pm 3.29 )</td>
</tr>
<tr>
<td>99.99%</td>
<td>( \pm 3.89 )</td>
</tr>
</tbody>
</table>

Confidence level & interval limits for \( z \) |

90%: 
\[
\begin{align*}
   & Pr(-k \leq z \leq k) = 0.90 \\
   & F(k) - F(-k) = 1 - \Phi(k) - [1 - \Phi(-k)] = \Phi(-k) - \Phi(k)
\end{align*}
\]

Recall: 
\[
   \Phi(-x) = 1 - \Phi(x)
\]

\[
   z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}
\]

\[
   \bar{X} = 0
\]

\[
   \sigma^2 = 1
\]

\[
   Q(k) = \frac{0.1}{2} = 0.05
\]

\[
   \Phi(k) = 0.90
\]

Look in App E "The Q-function" find what \( k \) if \( \Phi = 0.05 \). \( k \) is \( 0.1 \times 1.64 \) & 1.65 \( \rightarrow 1.645 \) 

The table above provides limits in \( z \) for different confidence levels. What are the limits for the sample mean \( \bar{X} \)?

\[
   -k \leq z \leq k \quad \Rightarrow \quad -k \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq k
\]

Multiply by \( \frac{\sigma}{\sqrt{n}} \)

\[
\begin{align*}
   & \bar{X} - k \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \bar{X} + k \frac{\sigma}{\sqrt{n}}
\end{align*}
\]

Lower limit for sample mean

Upper limit for sample mean

1) Confidence level \( \uparrow \) \( \rightarrow k \)
   \( \rightarrow \) Broader interval for sample mean (gain certainty, loose information)

2) \( \uparrow \) sample size \( n \) \( \rightarrow \) Narrower interval for sample mean
Cont. example 4.4.2:

For a very large population of voters with \( \mu = 100.5 \) (population mean) and \( s = 4.2 \) (sample st. dev.) we found the limits on \( \bar{z} \) for a confidence level of 95%, are \( \pm 1.96 \). For the sample mean \( \bar{X} \):

\[
\bar{X} - k \frac{s}{\sqrt{n}} \leq \bar{X} \leq \bar{X} + k \frac{s}{\sqrt{n}}
\]

\[
100 - 1.96 \times 4 \leq \bar{X} \leq 100 + 1.96 \times 4
\]

\[
92.16 \leq \bar{X} \leq 107.84
\]

Recall sample variance: \( s^2 = \frac{\sum x^2}{n} \) (large population) \( \Rightarrow \) sample st. dev. \( s = \sqrt{\frac{\sum x^2}{n}} \) = 4.2

What would be the interval limits for \( \bar{X} \) if the confidence level was 90% instead of 95%?

\[
100 - 1.645 \times 4 \leq \bar{X} \leq 100 + 1.645 \times 4
\]

\[
93.42 \leq \bar{X} \leq 106.58
\]

HW5: answers:

2.4/  a) \( \bar{X} = 0.2 \) ;  b) \( \bar{X} = 2p - 1 \) \( (p = \text{percentage in favor of candidate A}) \)

c) \( n = 10^6 \)

2.7/  a) \( p(117.6 \leq \bar{X} \leq 122.4) = 0.663 \)

b) \( p(117.6 \leq \bar{X} \leq 122.4) = 0.689 \)

3.3/  \( n = 822 \)

3.4/  Next page
\[ f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & x < 0 \end{cases} \]

How many samples do you estimate the variance of this random variable \( X \) with a std. dev. that is \( 5\% \) of the true value, using an unbiased estimator?

\[ \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad \rightarrow \quad \sqrt{\text{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}} \]

Sample Variance is a random variable

\[ \begin{align*}
\text{Mean of sample variance:} & \quad \bar{S}^2 = \frac{n-1}{n} \sigma^2 \\
\text{Unbiased:} & \quad \bar{S}^2 = \frac{n}{n-1} \bar{S}^2 \\
\text{Variance of sample variance:} & \quad \text{Var}(\bar{S}^2) = \frac{\mu_4 - \sigma^4}{n} \quad \text{(can be proved)}
\end{align*} \]

\[ \mu_4 = \frac{(X - \bar{X})^4}{n} \text{ is the fourth central moment of the population.} \]

\[ \text{Var}(\bar{S}^2) = n \frac{\mu_4 - \sigma^4}{(n-1)^2} \]

Correction:

\[ \frac{n}{(n-1)^2} (\mu_4 - \sigma^4) = 0.05^2 \sigma^4 \quad \rightarrow \quad (n^2 - 2n + 1)0.05^2 \sigma^4 = n(\mu_4 - \sigma^4) \]

\[ A = \frac{\mu_4 - \sigma^4}{0.05^2 \sigma^4} \quad \rightarrow \quad n^2 - (2A)n + 1 = 0 \]

\[ n = \frac{(2A) \pm \sqrt{(2A)^2 - 4}}{2} \]
\[ \sigma^2 = \text{second central moment of } X = \bar{X}^2 - \bar{X}^2 \]

\[ \bar{X} = \int_0^\infty dx \, e^{-x} x = \frac{\Gamma(2)}{1} = 2! = 2 \]

Recall: \[ \int_0^\infty dx \, x^n e^{-x} = \frac{n \Gamma(n+1)}{a^{n+1}} \]

\[ \Gamma(n+1) = n \Gamma(n) \]

\[ \bar{X}^2 = \int_0^\infty dx \, x^2 e^{-x} = \Gamma(3) = 3! = 6 \]

\[ \mu_4 \equiv (X - \bar{X})^4 = \int_0^\infty dx \, e^{-x} (x - 2)^4 = \cdots \]

\[ \frac{\text{eh}2 \pi/5.2 \text{ (Note: pg 65)}}{X^4 - 6 \sigma^2 \bar{X}^2 - \bar{X}^4 = 120 - 6 \times 2 \times 2 - 2^4} \]

\[ X^4 = \int_0^\infty dx \, e^{-x} x^4 = \Gamma(5) = 5! = 120 \]

\[ = 120 - 48 - 16 = 56 \]

\[ A = \frac{\mu_4 - \sigma^4}{\sigma^4} = \frac{56 - 4}{0.05 \times 2^4} = 5200 \Rightarrow n = \frac{5200 \pm \sqrt{5200^2 - 4}}{2} = 5202 \]

This is the sample size to estimate the sample variance whose standard deviation is 5\% of the true value.
Confidence levels and interval limits when \( n < 30 \):

Use Student t-distribution on variable \( T \) (analog of \( z \) when \( n > 30 \) using Gaussian distribution), with a degree of freedom,

\[
T = \frac{\bar{X} - \mu}{s / \sqrt{n}}
\]

\( d.o.f \quad \nu = n - 1 \)

Connection:

Same example 4.4.2 now for \( n < 30 \): Interval limits for sample mean when the confidence level was 95%:

\[
\Pr(-k_T \leq T \leq k_T) = 0.95
\]

\[
F_T(k_T) - F_T(-k_T) = 0.95
\]

Student t-distribution: property: \( F_T(-t) = 1 - F_T(t) \)

\( \nu = 8 \)

\[
2F_T(k_T) - 1 = 0.95 \Rightarrow F_T(k_T) = \frac{1.95}{2} = 0.975
\]

\( \text{App F: } \nu = 8 \Rightarrow k_T = 2.306 \)

Interval limits for \( T \) at 95% confidence level and \( \nu = 8 \) are \( \pm 2.306 \)
<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>$k_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.99%</td>
<td>$\pm 2.306$</td>
</tr>
<tr>
<td>99.1%</td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td></td>
</tr>
</tbody>
</table>

Internal limits for $\bar{X}$:

\[-k_T \leq \bar{X} - \bar{X} \leq k_T\]

\[-k_T \leq \frac{\bar{X} - \bar{X}}{\frac{S}{\sqrt{n}}} \leq k_T\]

\[\bar{X} - k_T \frac{S}{\sqrt{n}} \leq \bar{X} \leq \bar{X} + k_T \frac{S}{\sqrt{n}}\]

From example 4.4.2:

\[
\begin{align*}
\bar{X} &= 100 \Omega \\
S &= 4 \Omega \\
n &= 9
\end{align*}
\]

\[\bar{S} = \sqrt{n-1} s = \sqrt{\frac{9}{8}} 4\]

HW 6: (Ch4) 3.2 & 5.5 → due 4/23
Bipolar transistors made by 2 different companies. HYCAYN & ACW

Current gain as a random variable: 2 independent Gaussians with mean 120

Recall: Gaussian is defined when mean & variance (or std.) are specified.

From part 2.6: \[ \beta_H = 120; \sigma_H^2 = 100 \]
\[ \beta_A = 120; \sigma_A^2 = 25 \]

Mix of 20 & 20:
\[ \bar{\beta} = 120 \]
\[ \sigma^2 = \frac{\sigma_H^2 + \sigma_A^2}{2} = \frac{125}{2} = 62.5 \]

a) Ed selects a random sample of 10 (out of the 20 & 20 mix) with replacement, hence probability that sample mean \( \hat{\beta} \) is within 2\% of true mean \( \bar{\beta} \) is

\[ P(\beta(0.98) \leq \frac{10}{\bar{\beta}} \leq \beta(1.02)) = P(117.6 \leq \hat{\beta} \leq 122.4) = F(122.4) - F(117.6) = \phi\left(\frac{122.4 - 120}{\sqrt{62.5}}\right) - \phi\left(\frac{117 - 120}{\sqrt{62.5}}\right) \]

Recall: \[ F(x) = \phi\left(\frac{x - \mu}{\sigma_x}\right) \]
\[ \phi(-x) = 1 - \phi(x) \]
\[ F(0.96) - \phi(-0.96) = 2 \phi(0.96) - 1 = 2 \times 0.8315 - 1 = 0.663 \]

b) Repeat part a) if the sampling is without replacement: population is less than 40! need to be careful about \( \text{var} (\hat{\beta}) \): Recalling:

Recall: variance of sample mean:
\[ \text{var} (\hat{\mu}) = \frac{\sigma_x^2}{n} \frac{(N-n)}{(N-1)} \]

\[ n = \text{sample size} = 10 \]
\[ N = \text{population size} = \text{(less than 40 w/o replacement)} \]
\[ \text{var} (\hat{\mu}) = \frac{\sigma_x^2}{n} \text{ when } N \text{ is very large } (N \to \infty) \]
Sample of 10 out of 20 w/o replacement:

\[ \text{Var}(\hat{\beta}) = \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right) = \frac{6.25}{10} \left( \frac{40-10}{40-1} \right) \]

\[ = 4.807 \] (larger due to a smaller population w/o replacement)

\[ \rightarrow p_2 \left( 117.6 \leq \hat{\beta} \leq 122.4 \right) = \phi \left( \frac{122.4 - 120}{\sqrt{4.807}} \right) - \phi \left( \frac{117.6 - 120}{\sqrt{4.807}} \right) \]

\[ = \phi(1.095) - \phi(-1.095) \]

\[ = 2 \phi(1.095) - 1 \]

\[ = 2 \times 0.8621 - 1 = 0.7242 \]

---

HW5 Ch 4

7.4

Assign: \( X \) candidate

<table>
<thead>
<tr>
<th>1</th>
<th>( \rightarrow )</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>( \rightarrow )</td>
<td>B</td>
</tr>
</tbody>
</table>

1) Sample mean if

\[ \bar{X} = \frac{\sum_{i=1}^{n} f(x_i) x_i}{n} \quad \text{(more general def. of sample mean)} \]

Recall:

\[ \bar{X} = \frac{\sum_{i=1}^{n} \frac{1}{n} x_i}{n} \quad \text{(many } x_i\text{'s each with equal probability } \frac{1}{n} \text{)} \]

Here we have only two \( x_i\)'s a \( n=2 \), each has different probability.

\[ \bar{X} = \frac{2}{\sum_{i=1}^{2} f(x_i) x_i} = 0.6 \times 1 + 0.4 \times (-1) = 0.2 \]

5) Sample mean as a function of sample size \( n \) and percentage of people polled that prefers candidate A (in past a this percentage was 60%, but here we keep it as \( p_A \))

\[ \hat{X} = p_A \cdot 1 + (-p_A) \cdot (-1) = p_A - 1 \]
c) Sample size needed to estimate the sample mean with a standard deviation not greater than 0.1%:

\[ \text{Var} (\bar{X}) = \left( \frac{0.1}{100} \right)^2 \]

also for large population:

\[ \text{Var} (\bar{X}) = \frac{\sigma^2}{n} \]

\[ \left( \frac{0.1}{100} \right)^2 = \frac{1}{n} \rightarrow n = \frac{10000}{0.01} \]

\[ n = 10^6 \]

HW5 Ch4

(3.3) Similar to (3.4), except the distribution is uniform:

Random phase angle over a range of \( \pi \) \( \rightarrow f(\theta) = \frac{1}{2\pi} \) in the range of \( \pi \) outside.

Want to estimate variance of a sample of \( \theta \)\( \bar{X} \) is a random variable.

Find \( n \) such that the \( \text{stdev} \) of the sample variance is 0.05\( \sigma^2 \) using unbiased estimate

Variance of sample variance:

\[ \text{Var} (S^2) = \frac{\mu_4 - \sigma^4}{n} \]

Unbiased:

\[ S^2 \]

\[ \text{Var} (\bar{S}^2) = \frac{n \mu_4 - \sigma^4}{(n-1)^2} \]

Since:

\[ S^2 = \frac{n}{n-1} S \rightarrow \text{Var} (\bar{S}^2) = \frac{n-1}{n} \text{Var} (S^2) \]

As in problem 3.4 done earlier:

\[ A = \frac{\mu_4 - \sigma^4}{0.05 \sigma^4} \quad \text{and} \quad n^2 - (2A+1) = 0 \]

\[ n = \frac{(2+A)\pm\sqrt{(2A)^2-4}}{2} \]
In a uniform distribution \( U(-n, n) \):

\[
    f(x) = \begin{cases} \frac{1}{2n} & -n \leq x \leq n \\ 0 & \text{otherwise} \end{cases}
\]

\[
    \bar{x} = \frac{x_1 + x_2}{2} = 0; \quad \gamma_2 = \frac{(x_2 - x_1)^2}{12} = \frac{(2n)^2}{12} = \frac{n^2}{3}
\]

\[
    \mu_4 = (x - \bar{x})^4 = \int_{-n}^{n} dx \frac{1}{2n} x^4 = \frac{1}{2n} \left[ \frac{x^5}{5} \right]_{-n}^{n} = \frac{2n^5}{5n} = \frac{n^4}{5}
\]

\[
    \Rightarrow A = \frac{\mu_4 - \sigma^4}{0.05^2 \sigma^4} = \frac{n^4}{5} - \frac{n^4}{q} = \frac{4D^4}{45} - \frac{0.05^2 D^4}{q} = \frac{4}{5 \times 0.05^2} = 320
\]

\[
    \Rightarrow n = \frac{321 + \sqrt{321^2 - 4}}{2} = 314.99 \approx 322
\]
Hypothesis Testing:

- One-sided
- Two-sided

One-sided testing:

Manufacturer claims their capacitors have a breakdown voltage $V_{\text{breakdown}} \geq 300\, \text{V}$.

Dielectric insert breaks down at certain $V_{\text{applied}} = V_{\text{breakdown}}$.

We want higher $V_{\text{breakdown}}$ (more flexibility): we just need to test if the real or actual $V_{\text{breakdown}}$ is less than $300\, \text{V}$ – one-sided hypothesis testing.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Testing</th>
<th>Reject/Fail to Confirm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer claims $V_{\text{breakdown}} \geq 300, \text{V}$</td>
<td>$n = 100$, $\bar{X} = 290, \text{V}$, $S = 40, \text{V}$</td>
<td>99% Confidence Level</td>
</tr>
</tbody>
</table>

For $n > 30$, $\{\text{Gaussian}\}$, find interval limits for $z = \frac{X - \bar{X}}{\frac{S}{\sqrt{n}}}$.

- $z < z_\alpha$  $\Rightarrow$ Reject
- $z > z_\alpha$  $\Rightarrow$ Fail to reject

$p_\alpha (z \leq z_\alpha) = 0.99 = F(z_\alpha) = 1 - Q(z_\alpha)$

Recall $F(z) = \Phi \left( \frac{z - \bar{X}}{S} \right) = \Phi (z) = 1 - Q(z)$

$Q(z_\alpha) = 1 - 0.99 = 0.01$  $\Rightarrow$  $z_\alpha = \pm 2.33036$

Also:

- $z_\alpha = 2.33036$
- $Q(z_\alpha) = 0.01$  $\Rightarrow$  $X = 2.33036$
- $z_\alpha = 2.33036$

Graphs:

- Normal distribution
- $f(z)$
- $Q(z)$
- Asymptotic unit $z$
In our testing:

\[ z' = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{290 - 300}{\frac{40}{\sqrt{100}}} = \frac{-10}{4} = -2.5 \]

- Since \( z' \) is not in the interval of \( P(z > 2.1) = 0.99 \)
  - Reject hypothesis.

What can we change in the table to confirm the hypothesis?

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Testing</th>
<th>Reject or Confirm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer claims Breakdown &gt; 300 V</td>
<td>( \bar{X} = 290 V ) ( S = 40 V )</td>
<td>95% Confirm</td>
</tr>
</tbody>
</table>

\[ P(z < -2.1) = 0.05 = F(-2.1) = 1 - Q(2.1) \]
\[ Q(x) = 0.05 \text{ at } x = 1.64 \]
\[ Q(x) = 0.0495 \text{ at } x = 1.65 \]
\[ z = 1.64 \]

Interval is \( z > -1.64 \) for 95% confidence level.
- Our testing indicated \( z' = -2.5 \), falling outside the interval for 95%.
  - Reject

Summary

\[ z' = -2.5 \]

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Interval limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>( z &gt; -2.1 ) or ( z &gt; -2.32635 )</td>
</tr>
<tr>
<td>95%</td>
<td>( z &gt; -1.64 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Testing</th>
<th>Reject or Confirm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer claims Breakdown &gt; 300 V</td>
<td>( \bar{X} = 290 V ) ( S = 40 V )</td>
<td>99.9% Confirm</td>
</tr>
</tbody>
</table>
\[ P(z \leq z_1) = 0.999 \rightarrow Q(z_1) = 1 - 0.999 = 0.001 \rightarrow z_1 = 3.09023 \]

\[ \rightarrow \text{Interval is } z \geq -3.09 \text{ for } 99.9\% \text{ confidence level} \]

Since \( z' = -2.5 \) (from testing) belongs to this interval

**Two-sided Testing**

Manufacturer claims their ten diodes have Breakdown = 10V.

It doesn't matter whether Breakdown is larger or smaller.

Two-sided testing in this case.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Testing</th>
<th>Reject or Confirm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer claims [V_{\text{breakdown}} \leq 10\text{V} ]</td>
<td>( n = 100 ) [ \bar{X} = 10.3\text{V} ] [ S = 1.2\text{V} ]</td>
<td>Reject</td>
</tr>
</tbody>
</table>

\[ \text{Two-sided} \rightarrow P \left( -z_c \leq z \leq z_c \right) = 0.95 \rightarrow F(z_c) - F(-z_c) = 0.95 \]

\[ 1 - Q(z_c) = \left[ 1 - Q(-z_c) \right] \]

\[ 1 - Q(z_c) = 0.95 \rightarrow Q(z_c) = 0.05 \rightarrow \frac{z_c}{2} = 0.025 \]

\[ z_c = 1.96 \]

\[ \text{Interval limits for } z \text{ are } -1.96 \leq z \leq 1.96 \]

From our testing:

\[ z' = \frac{\bar{X} - X}{\frac{S}{\sqrt{n}}} = -1.2 \text{V} \]

\[ z' = \frac{10.3 - 10}{\frac{1.2}{10}} = 2.5 \text{ is outside this interval.} \]
Two-sided testing with small sample

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Testing</th>
<th>Reject or Confirm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer claims</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zener diode:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breakdown = 10 V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{n=9}{X} = 10.3$ V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\frac{S}{V} = 1.2$ V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence level 95%</td>
<td></td>
<td>Confirm</td>
</tr>
</tbody>
</table>

Gaussian $n > 30$
Student $n < 30$  ✓

Two-sided: $P( -t_c \leq t \leq t_c ) = 0.95$

$$F(t_c) - F(-t_c) = F(t_c) - [1 - F(t_c)] = (2F(t_c) - 1) = 0.95$$

$$\Rightarrow F(t_c) = \frac{1.95}{2} = 0.975$$

$$d.o.f = v = n - 1 = 8 \quad \Rightarrow \quad t_c = 2.306$$

From t-test: $t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{10.3 - 10}{1.2} = 0.75$ is inside the interval!

To confirm an hypothesis:
- Increase confidence level
- Decrease sample size
HW6 | Ch4:

3.2/ \( n = 5002 \)

5.5/ 9 capacitors: \( V_{\text{breakdown}} = 97, 104, 95, 98, 106, 94, 110, 103, 93 \) \( V \), \( n = 9 \)

a) \( \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = 99.78 \) \( V \)

b) Sample variance using unbiased estimate:
\[
S^2 = \frac{\left( \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1} \right)}{n-1} = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}
\]

\[
\hat{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2
\]

\[
\hat{S}^2 = \frac{1}{8} \left[ (99 - 99.78)^2 + \ldots + (93 - 99.78)^2 \right]
\]

\[
= 38.94
\]

c) 

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Testing</th>
<th>Reject or Confirm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer claims</td>
<td>( V_{\text{breakdown}} &gt; 100 )</td>
<td>( 95% ) confidence level</td>
</tr>
<tr>
<td>( n = 9 )</td>
<td>( \bar{X} = 99.78 ) ( V )</td>
<td>Confirm.</td>
</tr>
<tr>
<td>( S = 6.24 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( n = 9 \) \( \rightarrow \) Student's distribution.

One-sided:
\[
P_2 (t \leq t_c) = 0.95
\]
\[
F(t_c) = 0.95 \rightarrow t_c = 1.86
\]

Our interval is \( t \leq 1.86 \) or \( \boxed{-1.86 \leq t} \)

(Student's distribution is also symmetric: \( F(-t) = 1 - F(t) \))
From the testing \[ t' = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{99.78 - 100}{\frac{38.94}{\sqrt{9}}} = -0.017 \]

This falls in the interval $-1.68 \leq t' ! \rightarrow \text{Confirm}$

If it was two-sided: \[ 2F(t_c) - 1 = 0.95 \]
\[ F(t_c) = \frac{1.95}{2} = 0.975 \]

3.2) \( \bar{X} \text{ Gaussian } \rightarrow \bar{X} \sim N(0) \)

Sample: \( \bar{X} \) (sample mean)
\( S^2 \) (sample variance)

\[ \text{Var}(S^2) \] variance of sample variance.

\[ \Rightarrow \sqrt{\text{Var}(S^2)} = \text{std dev of sample variance} = 0.02 \cdot \sigma^2 \]

\( \Rightarrow \text{sample size } n ? \)

\[ \text{Var}(S^2) = \frac{n \mu_4 - \sigma^4}{n} \]
\( \mu_4 = \text{fourth central moment of the population} \)
\( \sigma^2 = \text{variance of prop.} \)

Unbiased estimate:

\[ S^2 = \frac{n}{n-1} \cdot \hat{S}^2 \]

\[ \Rightarrow \text{unbiased \hspace{1cm} biased} \]

\[ \text{Var}(S^2) = \left( \frac{n}{n-1} \right)^2 \cdot \text{Var}(\hat{S}^2) = \frac{n(n \mu_4 - \sigma^4)}{(n-1)^2} \]

\[ \Rightarrow \text{Var}(S^2) = \frac{n(n \mu_4 - \sigma^4)}{(n-1)^2} = \left[ 0.02 \cdot \sigma^2 \right]^2 \]

\[ \Rightarrow \text{solve for } n : \frac{n}{(n-1)^2} = \frac{0.02 \cdot \sigma^2}{\mu_4 - \sigma^4} = \frac{0.02 \cdot (10)^4}{30\cdot(10)^4} = \frac{0.02}{30} \\
\text{Gaussian with zero mean: } \mu_4 = 3\sigma^4 \]

\[ (n-1)(\mu_4 - \sigma^4) = 1.35^n \cdot (n-1) \sigma^4 \]
\[ n \text{ (even)} \quad n \text{ (odd)} \]
\[
\frac{n}{n^2 - 2n + 1} = 0.0002 \quad \rightarrow \quad B000 n = n^2 - 2n + 1
\]
\[
\rightarrow \quad n^2 - 5002 n + 1 = 0
\]
\[
n = \frac{5002 \pm \sqrt{5002^2 - 4}}{2}
\]
\[
\Rightarrow \quad n = 5002
\]

Chapter 5: Definition of Random Processes:

1) Continuous or discrete random processes

<table>
<thead>
<tr>
<th>Continuous</th>
<th>Discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example:</td>
<td>Example:</td>
</tr>
<tr>
<td>Breakdown in capacitors</td>
<td>Outcome of rolling a die</td>
</tr>
</tbody>
</table>

2) Non-deterministic & deterministic random processes

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Non-deterministic</th>
</tr>
</thead>
<tbody>
<tr>
<td>When a random variable is determined by another random variable: ( x(t) = A \cos(wt + \theta) ) where ( \theta ) is a random variable ( \rightarrow x ) is a deterministic random variable</td>
<td>When a random variable ( y(t) ) is not determined by any other random variable.</td>
</tr>
</tbody>
</table>

3) Stationary and non-stationary random processes

<table>
<thead>
<tr>
<th>Stationary</th>
<th>Non-stationary</th>
</tr>
</thead>
<tbody>
<tr>
<td>When the mean &amp; moments of that random variable are time independent (the variable itself can be time dependent)</td>
<td>Otherwise</td>
</tr>
</tbody>
</table>
4) **Ergodic and non-ergodic processes.**

<table>
<thead>
<tr>
<th>Ergodic</th>
<th>Non-ergodic</th>
</tr>
</thead>
<tbody>
<tr>
<td>If almost all members of an ensemble exhibit the same behavior (statistical mean, moments) as the whole population.</td>
<td>otherwise</td>
</tr>
<tr>
<td>It is possible to examine statistical behavior of whole population by examining only one member of the ensemble (or the sample).</td>
<td></td>
</tr>
</tbody>
</table>

If $X$ a random variable is both stationary & ergodic:

$$E[X]_{\text{statistical}} = \langle X(t) \rangle_{\text{time}}$$

<table>
<thead>
<tr>
<th>Ensemble average</th>
<th>Time average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[X] = \int_{-\infty}^{\infty} dx \ x f(x)$</td>
<td>$\lim_{T \to \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} dt \ x(t)$</td>
</tr>
</tbody>
</table>

Same for stationary & ergodic process.
Ch6: Correlation Functions

- Auto-correlation functions: $y$s of 2 copies of a same random variable
- Cross-correlation functions: $y$s of different random variables.

$X(t)$ random variable $\rightarrow \begin{cases} X_1 = X(t_1) & \text{random variable at time } t_1 \\ X_2 = X(t_2) & \text{time } t_2 \end{cases}$

- Auto-correlation function:
  \[
  R_X(t_1, t_2) = E[X_1 X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1) f(x_2) \,dx_1 \,dx_2
  \]

- If $X$ is stationary & ergodic:
  1) $X$ is stationary: stat. behavior (mean & moment) is time independent:

  \[
  R_X(t_1, t_2) = R_X(t_1 + T, t_2 + T) = R_X(0, t_2 - t_1) = R_X(\gamma)
  \]

  \[
  \gamma = \frac{t_2 - t_1}{2} \quad \text{(time difference)}
  \]

  \[
  \text{its auto-correlation function is only a function of the time difference}\ y$s the two copies of $X$.

  2) $X$ is also ergodic.

HW7 (Ch6): 1.2, 2.3, 3.2