HW 4 3.3:

X, Y are independent random variables. $\Leftrightarrow$ $f(x,y) = f_X(x) \cdot f_Y(y)$ — Density Function

Gaussian

$\bar{X} = 1 \quad \bar{Y} = 2$

$\sigma_X^2 = 1 \quad \sigma_Y^2 = 4$

$$Pr(XY > 0) = Pr(X \& Y < 0) + Pr(X \& Y > 0)$$

$$= Pr(X<0)\,Pr(Y<0) + Pr(X>0)\,Pr(Y>0)$$

$$= F_X(0)\; F_Y(0) + [1 - F_X(0)][1 - F_Y(0)]$$

$$= \underbrace{\phi\left(\frac{0-1}{1}\right)}_{\phi(-1)} \underbrace{\phi\left(\frac{0-2}{2}\right)}_{\phi(-1)} + \underbrace{[1-\phi(-1)]}_{\phi(1)}\underbrace{[1-\phi(-1)]}_{\phi(1)}$$

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

$$= [1 - 0.8413]^2 + (0.8413)^2 = 0.7329 \cong 0.733.$$

App. D pg 432

Recall: (Gaussian only.)

$$\boxed{F(x) = \phi\left(\frac{x - \bar{x}}{\sigma_x}\right) \quad \text{"phi"} \qquad \phi(-y) = 1 - \phi(y)}$$

$$\boxed{Pr(X < x) = F(x)}$$

Distrib. Function

4.4

Data

X	:	$\bar{X} = 0$	; $\sigma_X^2 = 1$
Y	:	$\bar{Y} = 0$	; $\sigma_Y^2 = 1$
Z	:	$\bar{Z} = 0$	; $\sigma_Z^2 = 1$

$X \& Y$ uncorrelated: $\rho_{XY} = 0$

$X \& Z$: $\rho_{XZ} = \frac{1}{2}$

$Y \& Z$: $\rho_{YZ} = -\frac{1}{2}$

a) $\sigma_W^2 = ?$

$$W = X + Y + Z \rightarrow \bar{W} = \bar{X} + \bar{Y} + \bar{Z} = 0$$

$$\sigma_W^2 = \overline{W^2} - \bar{W}^2 = E\{W^2\} = E[X^2 + Y^2 + Z^2 + 2XY + 2YZ + 2XZ]$$

$\bar{W} = 0$

$$= E[X^2] + E[Y^2] + E[Z^2] + 2E[XY] + 2E[YZ] + 2E[XZ]$$

$$= \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 + 2\underbrace{\rho_{XY}}_{0}\sigma_X\sigma_Y + 2\underbrace{\rho_{YZ}}_{-\frac{1}{2}}\sigma_Y\sigma_Z + 2\underbrace{\rho_{XZ}}_{\frac{1}{2}}\sigma_X\sigma_Z = 3\sigma_X^2 = 3$$

$$\boxed{\text{Recall } \rho_{XY} \equiv E\left[\frac{X-\bar{X}}{\sigma_X} \cdot \frac{Y-\bar{Y}}{\sigma_Y}\right] = E\left[\frac{XY}{\sigma_X\sigma_Y}\right]} \quad \bar{X} = \bar{Y} = 0$$

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

b) $\underset{\rho_{XW}\sigma_X\sigma_W}{\overset{\shortparallel}{E[XW]}} = E[X^2 + XY + XZ] = \underbrace{\sigma_X^2}_{1} + \cancel{1}\underbrace{\rho_{XY}}_{0}\sigma_X\sigma_Y + \underbrace{\rho_{XZ}}_{\frac{1}{2}}\sigma_X\sigma_Z = \frac{3}{2}$

Since $\rho_{XW} \equiv E\left[\frac{X-\bar{X}}{\sigma_X}\ \frac{W-\bar{W}}{\sigma_W}\right] = E\left[\frac{XW}{\sigma_X\sigma_W}\right]$

$$\rho_{XW} = \frac{3/2}{\sigma_X\sigma_W} = \frac{3/2}{1.\sqrt{3}} = \frac{\sqrt{3}}{2} \quad ✓$$

c) $\rho_{W(Y+Z)} = ?$

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

$$= \underbrace{\rho_{XY}}_{0}\sigma_X\sigma_Y + \underbrace{\rho_{XZ}}_{\frac{1}{2}}\sigma_X\sigma_Z + \sigma_Y^2 + 2\underbrace{\rho_{YZ}}_{-\frac{1}{2}}\sigma_Y\sigma_Z + \sigma_Z^2$$

$$\boxed{\rho_{W(Y+Z)}}\sigma_W\sigma_{(Y+Z)} = 2 - \frac{1}{2}\cdot1\cdot1 = \frac{3}{2}$$

$$\sigma_{(Y+Z)}^2 = \overline{(Y+Z)^2} - (\underbrace{\bar{Y}+\bar{Z}}_{0})^2$$

$$= \overline{Y^2} + \overline{Z^2} + 2\overline{YZ}$$

$$= \overline{Y^2} + \overline{Z^2} + \underbrace{2\rho_{YZ}\sigma_Y\sigma_Z}_{E[YZ] = \rho_{YZ}\sigma_Y\sigma_Z} = \sigma_Y^2 + \sigma_Z^2 + 2\rho_{YZ}\sigma_Y\sigma_Z = 2 + 2(-\tfrac{1}{2})1.1.$$

$$= 1$$

$$\rho_{W(Y+Z)} = \frac{3/2}{\sigma_W\sigma_{Y+Z}} = \frac{3/2}{\sqrt{3}\cdot 1} = \frac{\frac{3}{2}}{\sqrt{3}} = \frac{\sqrt{3}}{2} \quad ✓$$

Ch 4 Sampling Theory & Random Variables (Cont.)

Last time: → Mean of $\underbrace{\text{sample mean}}_{\text{a random variable}}$ $\Rightarrow$ $\overline{\hat{X}} = \overline{X}$ ↓ mean of population

(it varies from sample to sample)

→ Variance of sample mean: $\Rightarrow \text{Var}(\hat{X}) = \frac{\sigma_X^2}{n}\left(\frac{N-n}{N-1}\right) \xrightarrow{N\to\infty} \frac{\sigma_X^2}{n}$

n (size of a sample)
N (size of population)

→ Mean of $\underbrace{\text{sample variance}}$ $= \overline{S^2} = \frac{n-1}{n}\sigma_X^2$

random variable: varies from sample to sample

(n = size of a sample)

If we define the "unbiased" sample variance $\overline{\tilde{S}^2} \equiv \frac{n}{n-1}\overline{S^2}$

Then: $\boxed{\overline{\tilde{S}^2} = \sigma_X^2}$

mean of unbiased sample variance — variance of pop.

4.4: Sampling Distribution & Confidence Interval:

Example 4.4.2:

A very large population of resistors with $\overline{R} = 100\,\Omega$ (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 $= \hat{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

X_i: value of one resistor, a random variable,

$\begin{cases} n > 30 \to X_i \text{ follows a Gaussian distribution} \\ \quad Z \equiv \frac{\hat{X}-\overline{X}}{\sigma_X/\sqrt{n}} \\ n < 30 \to X_i \text{ follows a Student's T distribution} \\ \quad T \equiv \frac{\hat{X}-\overline{X}}{\tilde{S}} \end{cases}$

↓ degree of freedom $\nu = n-1$

HW5: Ch 4: 2.4; 2.7; 3.3; 3.4. (due 4/17) sample mean · pop. mean

$n > 30 \to$ x_i follows a Gaussian distribution

$$z \equiv \frac{\hat{\bar{X}} - \bar{X}}{\frac{\sigma_x}{\sqrt{n}}} \qquad \left(\text{Recall: } var(\hat{\bar{X}}) = \frac{\sigma_x^2}{n} \to \text{st. dev. of sample mean is } \frac{\sigma_x}{\sqrt{n}}\right)$$

$$\bar{z} = \frac{\overline{\hat{\bar{X}}} - \bar{X}}{\frac{\sigma_x}{\sqrt{n}}} = 0 \qquad \left(\text{Recall: } \overline{\hat{\bar{X}}} = \bar{X}\right)$$

$$var(X) = \overline{X^2} - \bar{X}^2$$

$$var(z) = \overline{z^2} - \bar{z}^2 = \overline{z^2} = \frac{\overline{(\hat{\bar{X}})^2 - 2\hat{\bar{X}}\bar{X} + \bar{X}^2}}{\left(\frac{\sigma_x}{\sqrt{n}}\right)^2}$$

$$= \frac{\overline{(\hat{\bar{X}})^2} - 2\overline{\hat{\bar{X}}}\bar{X} + \bar{X}^2}{\left(\frac{\sigma_x}{\sqrt{n}}\right)^2}$$

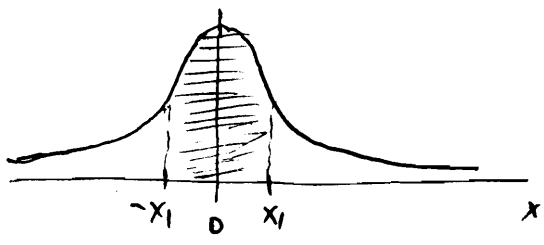
$$= \frac{\overline{(\hat{\bar{X}})^2} - \bar{X}^2}{\left(\frac{\sigma_x}{\sqrt{n}}\right)^2} = \frac{var(\hat{\bar{X}})}{\left(\frac{\sigma_x}{\sqrt{n}}\right)^2} = \frac{\frac{\sigma_x^2}{n}}{\frac{\sigma_x^2}{n}} = 1$$

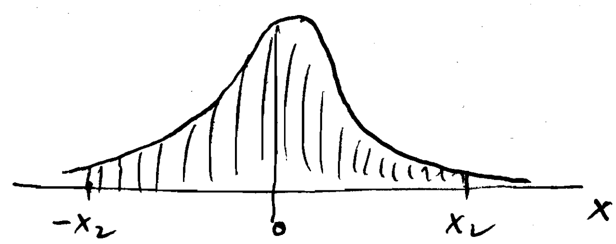
$$\boxed{\text{Recall: } var(\hat{\bar{X}}) = \frac{\sigma_x^2}{n} \text{ (large population)}}$$

$z \equiv \dfrac{\hat{\bar{X}} - \bar{X}}{\frac{\sigma_x}{\sqrt{n}}}$ 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:





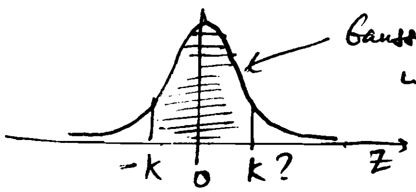
interval $(-x_1, x_1)$ is connected with a probability described by the area under the curve (shaded area)

interval $(-x_2, x_2)$ is connected with a larger probability (since $x_2 > x_1$) described by the area under the curve.

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 ($\bar{z} = 0$; $\sigma_z^2 = 1$) what are the limits $\pm k$ for the interval in z: $(-k, k)$?



Gaussian density function with $\bar{z} = 0$; $\sigma_z^2 = 1$

Two-sided test: $\underbrace{P_z(-k \le z \le k)}_{F_z(k) - F_z(-k)} = 0.95$ →

$$\phi(k) - \phi(-k) = 0.95$$
$$\phi(k) - (1 - \phi(k)) = 0.95$$
$$2\phi(k) - 1 = 0.95$$
$$\phi(k) = \frac{1.95}{2} = 0.975$$

$$\boxed{\text{Recall: Gaussian: } F(z) = \phi\left(\frac{z - \bar{z}}{\sigma_z}\right) = 1 - Q\left(\frac{z - \bar{z}}{\sigma_z}\right) \quad \phi(-x) = 1 - \phi(x)}$$

Look in App D : "Table for Normal Prob. Dist. Function Φ",
find what is k if $\Phi(k) = 0.975$. $\rightarrow k = 1.96$.

Confidence Level	Limits for interval in z $\pm k$
95%	± 1.96
90%	± 1.645
99%	? ± 2.58
99.9%	? ± 3.29
99.99%	? ± 3.89

Confidence level & interval limits for z

90% : $\Pr(-k \le z \le k) = 0.90$

$$F(k) - F(-k) = 1 - Q(k) - [1 - Q(-k)] = Q(-k) - Q(k)$$

$$\boxed{\begin{array}{l} F(z) = \Phi\left(\frac{z - \bar{z}}{\sigma_z}\right) = 1 - Q\left(\frac{z-\bar{z}}{\sigma_z}\right) = 1 - Q(z) \\ \bar{z} = 0 \\ \& \ \sigma_z^2 = 1 \\ \text{Recall: } Q(-x) = 1 - Q(x) \end{array}}$$

$$= 1 - Q(k) - Q(k)$$

$$= 1 - 2Q(k) = 0.90$$

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

$\rightarrow$ Look in App E "The Q-function" find what is k of $Q = 0.05$

$\rightarrow$ k is b/w 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 $\hat{X}$?

$$-k \le z \le k \quad \text{or} \quad -k \le \frac{\hat{X} - \bar{X}}{\frac{\sigma_X}{\sqrt{n}}} \le k$$

Multiply by $\frac{\sigma_X}{\sqrt{n}}$
Add $\bar{X}$

$$\rightarrow \underbrace{\bar{X} - k\frac{\sigma_X}{\sqrt{n}}}_{\substack{\text{lower limit} \\ \text{for sample mean}}} \le \hat{X} \le \underbrace{\bar{X} + k\frac{\sigma_X}{\sqrt{n}}}_{\substack{\text{upper limit} \\ \text{for sample mean}}}$$

1) Confidence level $\uparrow \rightarrow \uparrow k$
$\rightarrow$ broader interval for sample mean
(gain certainty, loose information)
2) $\uparrow$ sample size n $\rightarrow$ narrow the interval for $\hat{X}$

Cont. example 4.4.2:

For a very large population of resistors with $\bar{R} = 100\Omega$ (population mean) and $S = 4\Omega$ (sample ~~variance~~ st. dev.) we found the limits on z for a confidence level of 95% are ± 1.96. For the sample mean $\hat{X}$:

$$\bar{X} - k\frac{\sigma_x}{\sqrt{n}} \le \hat{X} \le \bar{X} + \frac{k\sigma_x}{\sqrt{n}}$$

$$100 - 1.96 \times 4 \le \hat{X} \le 100 + 1.96 \times 4$$

or

$$\boxed{92.16 \le \hat{X} \le 107.84}$$

Recall: sample variance: $S^2 = \frac{\sigma_x^2}{n}$ (large population) $\rightarrow$ sample st. dev $= \frac{\sigma_x}{\sqrt{n}} = 4\Omega$.

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

$$100 - 1.645 \times 4 \le \hat{X} \le 100 + 1.645 \times 4$$

$$\boxed{93.42 \le \hat{X} \le 106.58}$$

loose certainty (95% $\rightarrow$ 90%) gain information

HW5: answers:

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

c) $n = 10^6$

2.7/ a) $\Pr(117.6 \le \hat{\beta} \le 122.4) = 0.663$

b) $\Pr(117.6 \le \hat{\beta} \le 122.4) = 0.689$

3.3/ $n = 322$

3.4/ Next page

HW5 Ch4

3.4

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

<u>How many samples</u> to estimate the variance of this random variable $\bar{X}$ <u>w/a st. dev. that is 5% of the true value</u>, using an unbiased estimator.

$$\mathrm{Var}(\hat{\bar{X}}) = \frac{\sigma_x^2}{n} \rightarrow \sqrt{\mathrm{var}(\hat{\bar{X}})} = \frac{\sigma_x}{\sqrt{n}}$$

Sample Variance is a random Variable

<u>Mean</u> of sample variance: $\overline{S^2} = \frac{n-1}{n}\sigma_x^2$

<u>Unbiased</u>: $\rightarrow \overline{\tilde{S}^2} = \frac{n}{n-1}\overline{S^2} \rightarrow \sigma_x^2$ (large population)

Variance of sample variance: $\mathrm{Var}(S^2) = \frac{\mu_4 - \sigma^4}{n}$ (can be proved)

$\mu_4 \equiv \overline{(X-\bar{x})^4}$ is the fourth central moment of the population.

$$\boxed{\mathrm{Var}(\tilde{S}^2) = n\frac{\mu_4 - \sigma^4}{(n-1)^2}}$$

Correction:

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

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

$$\rightarrow n = \frac{(2+A) \pm \sqrt{(2+A)^2 - 4}}{2}$$

σ^2 = second central moment of X $= \overline{X^2} - \bar{X}^2$

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

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

$$\Gamma(n+1) = n\Gamma(n)$$

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

$\sigma^2 = 6 - 2^2 = 2$

$$\mu_4 \equiv \overline{(X-\bar{X})^4} = \begin{cases} = \int_0^\infty dx\, e^{-x} (x-2)^4 = \ldots \\ \text{ch2 pr. 5.2 (Notes pg 45)} \\ = \overline{X^4} - 6\sigma^2 \bar{X}^2 - \bar{X}^4 \end{cases} = 120 - 6\times 2\times 2^2 - 2^4$$

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

$$= 120 - 48 - 16 = 56$$

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

This is the sample size to estimate the sample variance whose standard deviation is 5% of the true value.

Ch 4 (Cont).

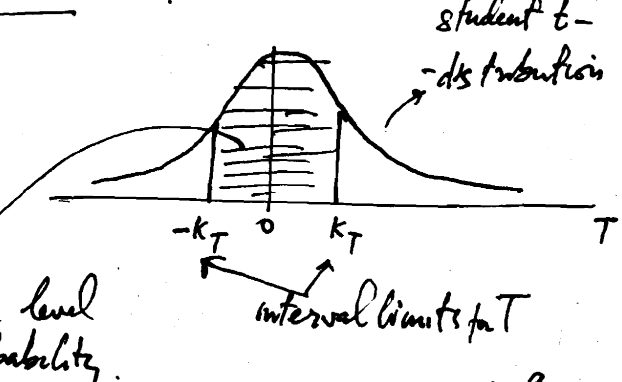
Confidence levels and interval limits when $n < 30$:

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

$$T \equiv \frac{\hat{\bar{X}} - \bar{X}}{\frac{\tilde{S}}{\sqrt{n}}}$$

d.o.f $\nu \equiv n-1$

→ Connection :



Same example 4.4.2 now for $n < 30$ ($n=9$): interval limits for sample mean when the confidence level was 95%:

$$Pr(-k_T \le T \le k_T) = 0.95$$

$$F_t(k_T) - F_t(-k_T) = 0.95$$

Student t-distribution : property: $F_t(-t) = 1 - F_t(t)$
$\downarrow$ table in App F (pg 436)
$\nu = 8$

$$\longrightarrow 2F_t(k_T) - 1 = 0.95 \Rightarrow F_t(k_T) = \frac{1.95}{2} = 0.975$$

$\downarrow$
App F : $\nu = 8 \rightarrow k_T = 2.306$

→ Interval limits for T at 95% confidence level and $\nu = 8$ are ± 2.306

Confidence Level	$\upsilon = 8$ k_T	$\upsilon = 20$ k_T
99.99%	.	
99.9%	.	
99%	.	
95%	± 2.306	± 2.086
90%	.	

Interval Limits for $\hat{X}$?

$$-k_T \leq T \leq k_T$$

$$-k_T \leq \frac{\hat{X} - \bar{X}}{\frac{\tilde{S}}{\sqrt{n}}} \leq k_T$$

$$\bar{X} - k_T \frac{\tilde{S}}{\sqrt{n}} \leq \hat{X} \leq \bar{X} + k_T \frac{\tilde{S}}{\sqrt{n}}$$

From example 4.4.2: $\left\{ \begin{array}{l} \bar{X} = 100\Omega \\ S = 4\Omega \\ n = 9 \end{array} \right. \rightarrow \boxed{\tilde{S} = \sqrt{\frac{n}{n-1}}\, S} = \sqrt{\frac{9}{8}}\, 4$

$\boxed{\text{HW6: (Ch4): } 3.2 \ \& \ 5.5 \longrightarrow \text{due } 4/23}$

HW5 Ch4

2.7 Bipolar transistors made by 2 different companies HYGAYN & ACE

↓

current gain → random variables: 2 : independent Gaussians with mean 120

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

From prob 2.6 :
$$\begin{cases} \overline{\beta_H} = 120 \; ; \; \sigma_H^2 = 100 \\ \overline{\beta_A} = 120 \; ; \; \sigma_A^2 = 25 \end{cases}$$

Mix of 20 & 20 :
$$\begin{cases} \overline{\beta} = 120 \\ \sigma^2 = \dfrac{\sigma_H^2 + \sigma_A^2}{2} = \dfrac{125}{2} = 62.5 \end{cases}$$

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

$\mathrm{var}(\hat{\beta}) = \frac{\sigma^2}{n} = \frac{\sigma^2}{10} = 6.25$

$$\Pr\left(\overline{\beta}(0.98) \le \hat{\beta} \le \overline{\beta}(1.02)\right) = \Pr\left(117.6 \le \hat{\beta} \le 122.4\right)$$

$$= F(122.4) - F(117.6) = \phi\left(\frac{122.4-120}{\sqrt{6.25}}\right) - \phi\left(\frac{117-120}{\sqrt{6.25}}\right)$$

$$\boxed{\text{Recall: } F(x) = \phi\left(\frac{x-\overline{x}}{\sigma_x}\right)}$$

$$\boxed{\phi(-x) = 1-\phi(x) \quad \text{App. D pg 432}}$$

$$= \phi(0.96) - \phi(-0.96)$$
$$= 2\phi(0.96) - 1$$
$$= 2 \times 0.8315 - 1$$
$$= 0.663. \checkmark$$

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

Recall: variance of a sample mean: $\boxed{\mathrm{var}(\hat{X}) = \frac{\sigma_X^2}{n}\left(\frac{N-n}{N-1}\right)}$

n = sample size (10)

N = population size (less than 40 w/o replacement).

$\mathrm{var}(\hat{X}) = \frac{\sigma_X^2}{n}$ when N is very large ($N \to \infty$)

Sample of 10 out of 20&20 mix w/o replacement:

$$Var(\hat{\beta}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{62.5}{10}\left(\frac{40-10}{40-1}\right)$$

$$= 4.807$$ (larger due to a smaller population w/o replacement)

$$\rightarrow Pr(117.6 \le \hat{\beta} \le 122.4) = \phi\left(\frac{122.4-120}{\sqrt{4.807}}\right) - \phi\left(\frac{117.6-120}{\sqrt{4.807}}\right)$$

$$= \Phi(1.095) - \Phi(-1.095)$$

$$= 2\,\Phi(1.095) - 1$$

$$= 2 \times 0.8621 - 1 = 0.7242$$

AppD.

HW5 Ch 4

7-4

assign: X		candidate
+1	$\longrightarrow$	A
−1	$\longrightarrow$	B

a) Sample mean if $\begin{cases} A \rightarrow 60\% \\ B \rightarrow 40\% \end{cases}$

$$\hat{X} = \sum_{i=1}^{n} f(x_i)\, x_i$$ (more general def. of sample mean).

Recall: $\hat{X} = \sum_{i=1}^{n} \frac{1}{n} x_i$ (many X_i's each with equal probability $\frac{1}{n}$)

Here we have only two X_i's or $n=2$, each has different probability.

$$\hat{X} = \sum_{i=1}^{2} f(x_i)\, x_i = 0.6 \times 1 + 0.4(-1) = 0.2$$

b) Sample mean as a function of sample size n and percentage of people polled that prefers candidate A (in part a this percentage was 60% but here we keep it as P_A) $\rightarrow$ $\hat{X} = P_A 1 + (1-P_A)(-1) = 2P_A - 1$

c) Sample size needed to estimate $\underbrace{\text{sample mean}}_{\text{a random variable}}$ with a standard deviation not greater than 0.1% :

$$Var(\hat{X}) = \left(\frac{0.1}{100}\right)^2 \sigma^2$$

also for large population: $Var(\hat{X}) = \frac{\sigma^2}{n}$

$$\left(\frac{0.1}{100}\right)^2 = \frac{1}{n}$$

$$\rightarrow n = \frac{10000}{0.01}$$

$$\boxed{n = 10^6}$$

HW5 Ch4

3.3 Similar to 3.4 except the distribution is uniform:

- Random phase angle over a range of 2π $\rightarrow$ $f(\theta) = \begin{cases} \frac{1}{2\pi} & \text{in the range of } 2\pi \\ 0 & \text{outside} \end{cases}$

- Want to estimate $\underbrace{\text{variance of a sample}}_{\text{is a random variable.}}$ of θ

Find n such that the st dev. of the sample variance is $0.05\sigma^2$ using unbiased estimate.

Variance of a sample variance: $Var(S^2) = \frac{\mu_4 - \sigma^4}{n}$

μ_4 = fourth central moment of the population

σ^2 = true variance of pop.

n = sample size

Unbiased: $\tilde{S}$

$$\boxed{Var(\tilde{S}^2) = n\frac{\mu_4 - \sigma^4}{(n-1)^2}}$$

Since: $\tilde{S}^2 \equiv \frac{n}{n-1} S^2 \rightarrow Var(\tilde{S}^2) = \frac{n^2}{(n-1)^2} Var(S^2)$

As in problem 3.4 done earlier: $A = \frac{\mu_4 - \sigma^4}{0.05^2\sigma^4}$ and $n^2 - (2+A) + 1 = 0$

or $n = \frac{(2+A) \pm \sqrt{(2+A)^2 - 4}}{2}$

In a uniform distribution b/w $-\pi$ & π : $f(x) = \begin{cases} \frac{1}{2\pi} & -\pi \leq x \leq \pi \\ 0 & \text{otherwise} \end{cases}$

$$\bar{X} = \frac{x_1 + x_2}{2} = 0 \quad ; \quad \sigma_X^2 = \frac{(x_2 - x_1)^2}{12} = \frac{(2\pi)^2}{12} = \frac{\pi^2}{3}$$

$$\mu_4 = \overline{(X - \bar{X})^4} = \int_{-\pi}^{\pi} dx \ \frac{1}{2\pi} x^4 = \frac{1}{2\pi} \left[\frac{x^5}{5} \right]_{-\pi}^{\pi} = \frac{2\pi^5}{2\pi \cdot 5} = \frac{\pi^4}{5}$$

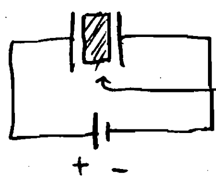
$$\rightarrow A = \frac{\mu_4 - \sigma^4}{0.05^2 \sigma^4} = \frac{\frac{\pi^4}{5} - \frac{\pi^4}{9}}{0.05^2 \frac{\pi^4}{9}} = \frac{\frac{4\cancel{\pi^4}}{45}}{0.05^2 \frac{\cancel{\pi^4}}{9}} = \frac{4}{5 \times 0.05^2} = 320$$

$$\rightarrow n = \frac{322 \pm \sqrt{322^2 - 4}}{2} = 319.99 \approx \boxed{322}$$

Hypothesis Testing: $\begin{cases} \text{One-sided} \\ \text{Two-sided} \end{cases}$

One-sided testing:

Manufacturer claims their capacitors have $\underbrace{V_{Breakdown}} \geq 300\,V$



dielectric insert: breaks down at certain $V_{Applied} = V_{Breakdown}$

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

Hypothesis	Testing	Reject or Confirm
Manufacturer claims $V_{Breakdown} \geq 300V$	$n = 100$ $\hat{X} = 290V$ $\tilde{S} = 40V$	Confidence level 99% Reject

↓ $n > 30$ → $\begin{cases} \text{Gaussian} ✓ \\ \text{Student-t} \end{cases}$ → Find interval limits for $z \equiv \dfrac{\hat{X} - \bar{X}}{\frac{\sigma}{\sqrt{n}}}$

$\hookrightarrow \begin{Bmatrix} \bar{z} = 0 \\ \sigma_z^2 = 1 \end{Bmatrix}$

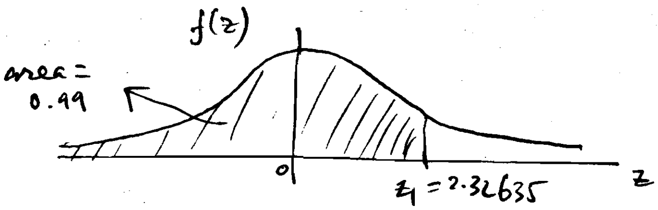
$$P_r(z \leq z_1) = 0.99 = F(z_1) = 1 - Q(z_1)$$

Gaussian

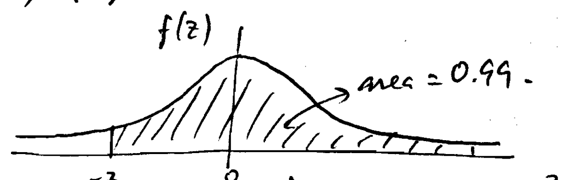
$\hookrightarrow$ Recall $F(z) \equiv \Phi\left(\dfrac{z - \bar{z}}{\sigma_z}\right) = \Phi(z) = 1 - Q(z)$

$\hookrightarrow Q(z_1) = 1 - 0.99 = 0.01 \rightarrow \boxed{z_1 = 2.32635}$

App E $\hookrightarrow Q(x) = 10^{-2} \rightarrow x = 2.32635$



also ↓ Gaussian: symmetric wrt $\bar{z}$



↓ $P_r(z \geq -z_1) = 0.99$ at

In our testing: $z' = \dfrac{\hat{\bar{X}} - \bar{X}}{\frac{\sigma}{\sqrt{n}}} = \dfrac{\overset{\text{Testing}}{240} - \overset{\text{Claim}}{300}}{\frac{40}{\sqrt{100}}} = \dfrac{-10}{4} = -2.5$

→ Since z' is not in the interval of $P_r(z \geq z_1) = 0.99$

→ Reject hypothesis.

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

Hypothesis	Testing	Reject or Confirm
Manufacturer claims $V_{Breakdown} \geq 300V$	$n = 100$ $\hat{\bar{X}} = 290V$ $\tilde{S} = 40V$	Confidence Level 95% Reject

$$P_r(z \leq z_1) = 0.95 = F(z_1) = 1 - Q(z_1)$$

$$\rightarrow Q(z_1) = 1 - 0.95 = 0.05 \rightarrow \begin{cases} Q(x) = 0.0505 \rightarrow x = 1.64 \\ Q(x) = 0.0495 \rightarrow x = 1.65 \end{cases}$$

App E

$\boxed{z_1 = 1.64}$

Interval is $z \geq -1.64$ for 95% confidence level.

→ our testing indicates $z' = -2.5$, falling outside interval for 95%

→ Reject

Summary

$z' = -2.5$ (from Testing)

Confidence level	Interval limits
99%	$z \geq -z_1$ or $z \geq -2.32635$
95%	$z \geq -z_1$ or $z \geq -1.64$

Hypothesis	Testing	Reject or Confirm
Manufacturer claims $V_{Breakdown} \geq 300V$	$n=100$ $\hat{\bar{X}} = 290V$ $\tilde{S} = 40V$	Confidence level: 99.9% Confirm.

$Pr(z \le z_1) = 0.999 \rightarrow Q(z_1) = 1 - 0.999 = 0.001 \xrightarrow[\text{AppE}]{} z_1 = 3.09023$

$\rightarrow$ Interval is $z \geq -3.09$ for 99.9 % confidence level
Since $z' = -2.5$ (from testing) belongs to this interval $\rightarrow$ Confirm.

Two-sided Testing :

Manufacturer claims their Zener diodes have $V_{Breakdown} \simeq 10V$
It doesn't matter whether $V_{Breakdown}$ is larger or smaller. $\rightarrow$
Two-sided testing in this case.

Hypothesis	Testing	Reject or Confirm Confidence level at 95%
Manufacturer claims $V_{Breakdown} \approx 10V$	$n = 100$ $\hat{X} = 10.3V$ $\tilde{S} = 1.2V$	Reject.

$\begin{cases} \text{Gaussian} & n > 30 \quad \checkmark \\ \text{Student } t & n < 30 \end{cases}$

Two-sided $\rightarrow$ $Pr(-z_c \le z \le z_c) = 0.95 \rightarrow F(z_c) - F(-z_c) = 0.95$

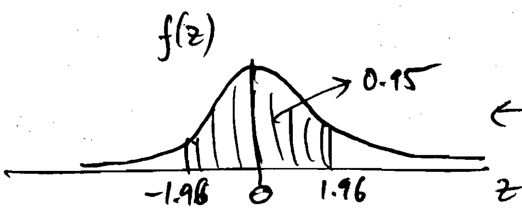
$1 - Q(z_c) - [1 - Q(-z_c)]$

$\boxed{\text{Recall: } Q(-x) = 1 - Q(x)}$

$1 - Q(z_c) - [1 - (1 - Q(z_c))]$

$1 - 2Q(z_c) = 0.95$

$Q(z_c) = \frac{0.05}{2} = 0.025.$



$\rightarrow$ AppE $z_c = 1.96$

$\rightarrow$ Interval limits for z are $\boxed{-1.96 \le z \le 1.96}$

From our testing : $z' = \dfrac{\hat{X} - \bar{X}}{\frac{\sigma}{\sqrt{n}}} = \dfrac{10.3 - 10}{\frac{1.2}{10}} = 2.5$ is outside this interval.

$\rightarrow$ Reject

Two-sided testing with small sample:

Hypothesis	Testing	Reject or Confirm
Manufacturer claims Zener diode: $V_{breakdown} \simeq 10V$	$n = 9$ $\hat{X} = 10.3V$ $\tilde{S} = 1.2V$	Confidence level 95% Confirm

Gaussian $n > 30$
Student-t $n < 30$ ✓

Two-sided. $Pr(-t_c \le t \le t_c) = 0.95$

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

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

$$d.o.f = \nu \equiv n - 1 = 8 \xrightarrow[\text{Appendix F}]{} t_c = 2.306$$

Interval limits for t are $-2.306 \le t \le 2.306$. (95%)

From our testing $t' = \dfrac{\hat{X} - \bar{X}}{\frac{\sigma}{\sqrt{n}}} = \dfrac{10.3 - 10}{\frac{1.2}{\sqrt{9}}} = 0.75$ is inside the interval!

→ Confirm.

→ To confirm an hypothesis : $\left\{ \begin{array}{l} \text{increase confidence level} \\ \text{decrease sample size.} \end{array} \right.$

→ Ex 2 on 4/29 (on Ch 3 & 4 and HW's 4,5 & 6)

HW6 / Ch4:

3.2/ $n = 5002$

5.5/ 9 capacitors: $V_{Breakdown}$: 97, 104, 95, 98, 106, 92, 110, 103, 93 V

$$n = 9$$

a) $$\hat{X} = \frac{\sum_{i=1}^{9} x_i}{9} = 99.78 \text{ V}$$

b) Sample variance using unbiased estimate:

$$S^2 = \overline{\hat{X}^2} - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{X})^2 \qquad \tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{X})^2$$

$$\tilde{S}^2 = \frac{n}{n-1} S^2$$

$$\rightarrow \tilde{S}^2 = \frac{1}{8}\left[(97 - 99.78)^2 + \cdots \cdots (93 - 99.78)^2\right]$$

$$= 38.94.$$

c)

Hypothesis	Testing	Reject or Confirm
Manufacturer claims capacitors $V_{Breakdown} \geq 100V$	ONE-SIDED. $n=9$, $\hat{X} = 99.78V$	95% confidence level. Confirm.

$\tilde{S} = 6.24 \text{ V}$

$n = 9 \rightarrow$ Student't distribution.

One-sided: $Pr(t \leq t_c) = 0.95$

$F(t_c) = 0.95 \rightarrow t_c = 1.86$ ↓ AppF

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

(Student-t distribution is also symmetric: $F(-t) = 1 - F(t)$)

From the testing $\quad t' = \dfrac{\hat{X} - \overline{X}}{\frac{\tilde{S}}{\sqrt{n}}} = \dfrac{99.78 - 100}{\frac{38.94}{\sqrt{9}}} = -0.017$

This falls in the interval $-1.86 \le t$! $\rightarrow$ $\boxed{\text{Confirm}}$

---

if it was two-sided:

$$2F(t_c) - 1 = 0.95$$
$$F(t_c) = \frac{1.95}{2} = 0.975$$

3.2/ $\quad X$ Gaussian $\rightarrow \overline{X} = 0$

Samples $\rightarrow \hat{X}$ (sample mean)

S^2 (sample variance)

$\operatorname{var}(S^2)$ variance of sample variance.

$\hookrightarrow \boxed{\sqrt{\operatorname{var}(\tilde{S}^2)} \; : \text{st dev of sample variance} := 0.02\,\sigma^2}$

$\hookrightarrow$ sample size n?

$$\operatorname{Var}(S^2) = \frac{\mu_4 - \sigma^4}{n}$$

μ_4 = fourth central moment of the population

σ^2 = variance of pop.

Unbiased estimate: $\quad \underbrace{\tilde{S}^2}_{\text{unbiased}} = \dfrac{n}{n-1} \underbrace{S^2}_{\text{biased}}$

$$\boxed{\operatorname{Var}(\tilde{S}^2) = \left(\frac{n}{n-1}\right)^2 \operatorname{Var}(S^2) = \frac{n(\mu_4 - \sigma^4)}{(n-1)^2}}$$

$$\rightarrow \operatorname{var}(\tilde{S}^2) = \frac{n(\mu_4 - \sigma^4)}{(n-1)^2} = \left[0.02\,\sigma^2\right]^2$$

$\rightarrow$ Solve for n: $\quad \dfrac{n}{(n-1)^2} = \dfrac{0.02^2\sigma^4}{\mu_4 - \sigma^4} = \dfrac{0.02^2\sigma^4}{3\sigma^4 - \sigma^4} = \overset{0.0002}{0.01}$

Gaussian w/ zero mean: $\mu_4 = 3\sigma^4 \quad \left(\overline{(X - \overline{X})^n} = \begin{cases} 1.3.5\ldots(n-1)\,\sigma_x^n & n \text{ (even)} \\ 0 & n \text{ (odd)} \end{cases} \right)$

$$\frac{n}{n^2-2n+1} = 0.0002 \rightarrow 5000\,n = n^2 - 2n + 1$$

$$\rightarrow n^2 - 5002\,n + 1 = 0$$

$$n = \frac{5002 \pm \sqrt{5002^2 - 4}}{2}$$

$$\Rightarrow \boxed{n = 5002}$$

Ch 5 Definition of Random Processes:

1) Continuous or discrete random processes

	Continuous	Discrete
Example:	V Breakdown for capacitors	Outcomes of rolling a dice

2) Non-deterministic & deterministic random processes

Deterministic	Non-deterministic
When a random variable is determined by another random variable: $X(t) = A\cos(\omega t + \theta)$ where θ is a random variable → X is a deterministic random variable	When a random variable $y(t)$ is not determined by any other random variable.

3) Stationary and non-stationary random process.

Stationary	Non-stationary
When the mean & moments of that random variable are time independent (the variable itself can be time dependent)	Otherwise

4) Ergodic and non-ergodic process.

Ergodic	Non-ergodic
→ If almost all members of an ensemble exhibits the same behavior (statistical: mean, moments) as the whole population → It is possible to examine stat. behavior of whole population by examining only a member of the ensemble (or one sample)	otherwise

If $\underset{\sim}{X}$ a random variable is both stationary & ergodic

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

↓ ensemble average = stat. average using density function

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

↓ time average :

$$\boxed{\lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} dt \; x(t)}$$

same for stationary & ergodic process.

Ch6: Correlation Functions

→ Auto-correlation functions: b/w 2 copies of a same random variable

→ Cross-correlation functions: b/w different random variable.

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

↳ Auto-correlation function: $R_X(t_1, t_2) = E[X_1 X_2]$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_1\, dx_2\, x_1 x_2\, f(x_1) f(x_2)$$

↳ If X is stationary & ergodic:

1) X is stationary: stat. behavior (mean & moments) is time independent:

$$R_X(t_1, t_2) = R_X(t_1 + T, t_2 + T)$$

$$T = -t_1 \rightarrow R_X(t_1, t_2) = R_X(0, \underbrace{t_2 - t_1}_{\tau \text{ (time difference)}}) = R_X(\tau)$$

↳ its autocorrelation function is only a function of the time difference b/w the two copies of X.

2) X is also ergodic.

HW7 (Ch6): 1.2; 2.3; 3.2