HW #2 Quality Control, Spring 2013, [+10 points]
Due 4 / 9 (Tue) [Submit after the Tuesday seminar, or put in my mailbox]

Q1. Range method [+3]

Let $X_1, X_2 \overset{iid}{\sim} N(\mu, \sigma^2)$ and $W = \dfrac{X_{(2)} - X_{(1)}}{\sigma}$ be the relative range.

1) Find the p.d.f. of W.

2) Calculate $d_2 = E[W]$ and compare it with Table's value.

3) Do you know the name of this distribution of W ?

(answer is not unique; clearly state why do you call its name)

Q2. ARL [+2]

The average run length is **numerically** obtained by the Monte Carlo method as follows:

$$E[L] \approx \frac{1}{M}\sum_{i=1}^{M} L_i, \quad M = 30000,$$

where $L_1, \ldots, L_M$ are iid replications of L.

1) Calculate the in-control ARL under the 3-sigma limits by Monte Carlo simulation (with codes). Compare your result with the **theoretical** value.

2) Do the same thing as 1) for $SD[L]$.

Q3. P-chart [+2] A manufacture conducts a waterproof testing for 5 electric boards. The number of defective circuits on the board is recorded as follows:

Board ID	1	2	3	4	5
The number of defectives circuits	20	20	23	25	12
The number of circuits	200	200	200	200	200

1. Draw a p-control chart.
2. The p-chart based on 3-sigma limits relies on the normal approximation. Please discuss whether the normal approximation to this dataset is suitable or not.

(answer is not unique; give numerical or theoretical evidence for your answer)

Q4. OC curve for P-chart [+2]

Draw OC curve for n=50 with LCL=0.0303, UCL=0.3697 by directly computing the Binomial probability masses (using computer). Compare your results with the OC curve based on hand writing.

Quality Control QW #2

+10/10

Q1. +3

1) We have $X_1, X_2 \overset{iid}{\sim} N(\mu, \sigma^2)$

Then $\frac{X_1-\mu}{\sigma}, \frac{X_2-\mu}{\sigma} \overset{iid}{\sim} N(0,1)$

And $\max\{\frac{X_1-\mu}{\sigma}, \frac{X_2-\mu}{\sigma}\} = \frac{X_{(2)}-\mu}{\sigma}$, $\min\{\frac{X_1-\mu}{\sigma}, \frac{X_2-\mu}{\sigma}\} = \frac{X_{(1)}-\mu}{\sigma}$

Let $U = \frac{X_{(1)}-\mu}{\sigma}$ and $V = \frac{X_{(2)}-\mu}{\sigma}$

➔ The joint pdf of U,V is

$$f_{U,V}(u,v) = 2![\phi(x)]^2 = 2\frac{1}{2\pi}e^{-\frac{u^2+v^2}{2}} = \frac{1}{\pi}e^{-\frac{u^2+v^2}{2}} \quad , -\infty<u<v<\infty$$ ✓

Again, let Y=U and W=V-U

⇔ U=Y , V=Y+W ➔ The Jacobian is $J = \begin{vmatrix} \frac{\partial U}{\partial Y} & \frac{\partial U}{\partial W} \\ \frac{\partial V}{\partial Y} & \frac{\partial V}{\partial W} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$

➔ the joint pdf of Y,W is

$$f_{Y,W}(y,w) = f_{U,V}(y, y+w)|1| = \frac{1}{\pi}e^{-\frac{y^2+(y+w)^2}{2}} = \frac{1}{\pi}e^{-\frac{1}{2}(y^2+y^2+2yw+w^2)}$$

$$= \frac{1}{\pi}e^{-\frac{1}{2}w^2}e^{-(y^2+wy)} = \frac{1}{\pi}e^{-(\frac{1}{2}w^2-(\frac{w}{2})^2)}e^{-(y^2+wy+(\frac{w}{2})^2)}$$

$$= \frac{1}{\pi}e^{-\frac{w^2}{4}}e^{-(y+\frac{w}{2})^2} \quad , -\infty<y<\infty , w>0$$

So the marginal pdf of W is

$$f_W(w) = \int_{-\infty}^{\infty} f_{Y,W}(y,w)dy = \int_{-\infty}^{\infty}\frac{1}{\pi}e^{-\frac{w^2}{4}}e^{-\left(y+\frac{w}{2}\right)^2}dy = \frac{1}{\pi}e^{-\frac{w^2}{4}}\int_{-\infty}^{\infty}e^{-(y+\frac{w}{2})^2}dy$$

$$= \frac{1}{\pi}e^{-\frac{w^2}{4}}\int_{-\infty}^{\infty}e^{-(y+\frac{w}{2})^2}d(y+\frac{w}{2}) = \frac{1}{\pi}e^{-\frac{w^2}{4}}\sqrt{\pi}$$

✓ $$= \frac{1}{\sqrt{\pi}}e^{-\frac{w^2}{4}} \quad , w>0$$

Let $I = \int_{-\infty}^{\infty} e^{-x^2}dx$

$\Rightarrow I^2 = \int_{-\infty}^{\infty} e^{-x^2}dx \int_{-\infty}^{\infty} e^{-y^2}dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}dx\,dy$

$= \int_0^{\infty}\int_0^{2\pi} e^{-r^2}(r\,d\theta\,dr) = \int_0^{\infty} re^{-r^2}\int_0^{2\pi}d\theta\,dr$

$= 2\pi\int_0^{\infty} re^{-r^2}dr = 2\pi(-\frac{1}{2}e^{-r^2}|_0^{\infty})$

$= \pi$

$\Rightarrow I = \sqrt{\pi}$

2)

d_2=E[W]=$\int_0^\infty w\frac{1}{\sqrt{\pi}}e^{-\frac{w^2}{4}}dw$=$\frac{1}{\sqrt{\pi}}\int_0^\infty w\,e^{-\frac{w^2}{4}}dw$=$\frac{1}{\sqrt{\pi}}(-2e^{-\frac{w^2}{4}}\Big|_0^\infty)$= $\frac{1}{\sqrt{\pi}}2 \fallingdotseq 1.128$

The value of d_2 table at n=2 is1.128

So we find that the result is similar to the table's value.

3)

We have $f_W(w)=\frac{1}{\sqrt{\pi}}e^{-\frac{w^2}{4}}$, w>0

And we can rewrite it as $f_W(w)=2\frac{1}{2\sqrt{\pi}}e^{-\frac{w^2}{2\times 2}}$, w>0

It's similar to the pdf of N(0,2).

But W is with positive support and its pdf is twice as much as the pdf of N(0,2).

So I call it as "positive normal". (Because I think "twice" is used to adjust the pdf such that its integration in (0, ∞) equal to 1)

Q2. +2/2

1) We have X_{ij}, i=1,…,m , j=1,…,n $\overset{iid}{\sim}$ N(μ,σ^2)

$\Rightarrow \overline{X_{i\cdot}}$, i=1,…,m $\overset{iid}{\sim}$ N $(\mu,\frac{\sigma^2}{n})$

$\Rightarrow \frac{\overline{X_{i\cdot}}-\mu}{\sigma/\sqrt{n}} \overset{iid}{\sim}$ N(0,1)

And we say that process is out-of-control if $\overline{X_{i\cdot}}$>μ_0+3$\frac{\sigma}{\sqrt{n}}$ or $\overline{X_{i\cdot}}$<μ_0-3$\frac{\sigma}{\sqrt{n}}$

$\Leftrightarrow \frac{\overline{X_{i\cdot}}-\mu_0}{\sigma/\sqrt{n}}>3$ or $\frac{\overline{X_{i\cdot}}-\mu_0}{\sigma/\sqrt{n}}<-3$

So generate some random variet from N(0,1) to simulate in-control ARL

```
> run_length=function(n)#use n-sigma
+ {
+ k=1
+ repeat{
+ x=rnorm(1)
+ if(abs(x)>n)break
+ k=k+1
+ }
+ return(k)
+ }
>
> L=numeric()#generate Li
> for(i in 1:30000){
+ temp=run_length(3)
+ L[i]=temp
+ }
> mean(L)#average Li
[1] 372.1721
```

And the theoretical value of in-control ARL is

$$E[L\,|\,\mu=\mu_0]=\frac{1}{P(\overline{X_{i\cdot}}>\mu0+3\frac{\sigma}{\sqrt{n}} \text{ or } \overline{X_{i\cdot}}<\mu0-3\frac{\sigma}{\sqrt{n}}|\mu=\mu_0)}=\frac{1}{P(\frac{\overline{X_{i\cdot}}-\mu_0}{\sigma/\sqrt{n}}>3 \text{ or } \frac{\overline{X_{i\cdot}}-\mu_0}{\sigma/\sqrt{n}}<-3|\mu=\mu_0)}$$

$$=\frac{1}{1-\Phi(3)+\Phi(-3)}\fallingdotseq\frac{1}{0.0027}\fallingdotseq 370$$

So the two values are close.

2)

We use the same data from 1)

And then calculate the standard deviation of the data

```
> sqrt(var(L))#SD of Li's
[1] 371.5853
```

Also, the theoretical value of $SD_0[L]$ is

$$SD_0[L]=\frac{\sqrt{1-p}}{p}$$

where $p=P(\frac{\overline{X_{i\cdot}}-\mu_0}{\sigma/\sqrt{n}} > 3 \text{ or } \frac{\overline{X_{i\cdot}}-\mu_0}{\sigma/\sqrt{n}} < -3|\mu = \mu_0)=1-\Phi(3)+\Phi(-3)=0.0027$

⇨ the theoretical value of $SD_0[L]$ is $\frac{\sqrt{1-0.0027}}{0.0027} \fallingdotseq 369.87$

We find that two values are close.

Q3.

+2/2

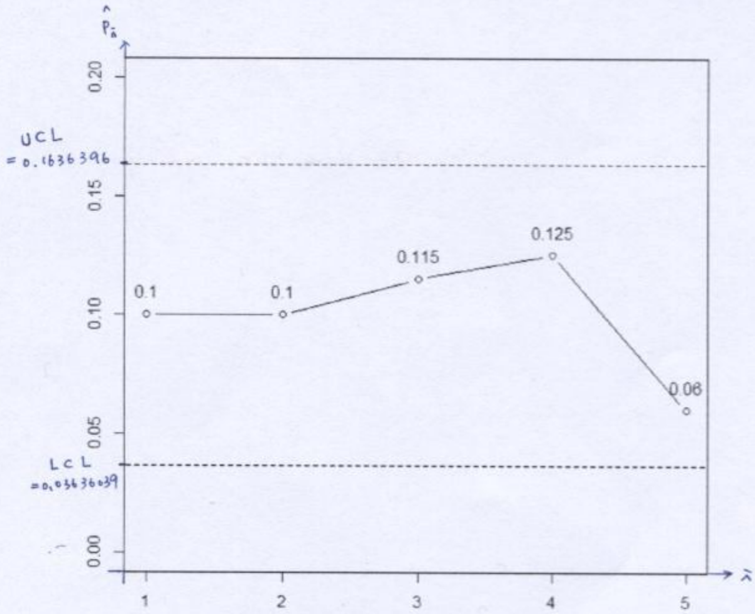
1.

n=200	D_i	$\widehat{p_i}$
1	20	0.1
2	20	0.1
3	23	0.115
4	25	0.125
5	12	0.06

⇨ $\bar{p}$=0.1

$$UCL=\bar{p}+3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}=0.1+3\sqrt{\frac{0.1(1-0.1)}{200}}\fallingdotseq 0.1636396$$

$$LCL=\bar{p}-3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}=0.1-3\sqrt{\frac{0.1(1-0.1)}{200}}\fallingdotseq 0.03636039$$



Reference ?

2. In the course of biostatistics, many theorem or lemma of binomial distribution which use normal approximation are requaired $n \cdot p \cdot (1-p) \geq 5$.

So I calculate $n \cdot p \cdot (1-p)$ in this case and the result is $200 \times 0.1 \times 0.9 = 18$

18 is larger than 5.

So I think the normal approximation is suitable in this case.

Q4. +2/2

We have $\beta=P(LCL \le \widehat{p_i} \le UCL \mid H_1: p \ne p_0) = P(n \times LCL \le D_i \le n \times UCL \mid H_1: p \ne p_0)$

$$=\sum_{x=[n\times LCL]+1}^{[n\times UCL]} \binom{n}{x} p^x (1-p)^{n-x}$$

Here, n=50

$[n\times UCL]=[50\times 0.3697]=[18.485]=18$

$[n\times LCL]=[50*0.0303]=[1.515]=1$

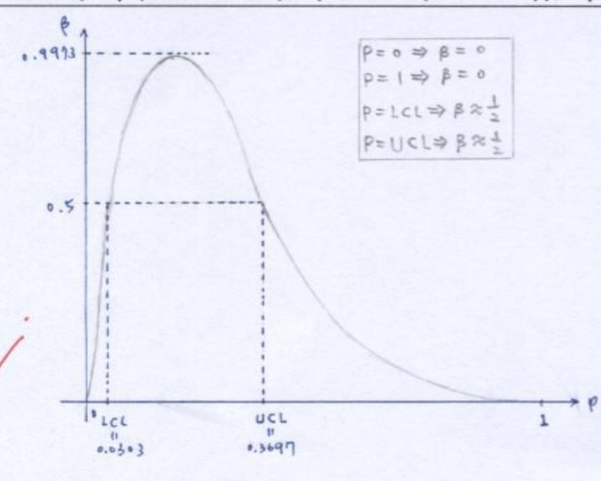
So make a code as following.

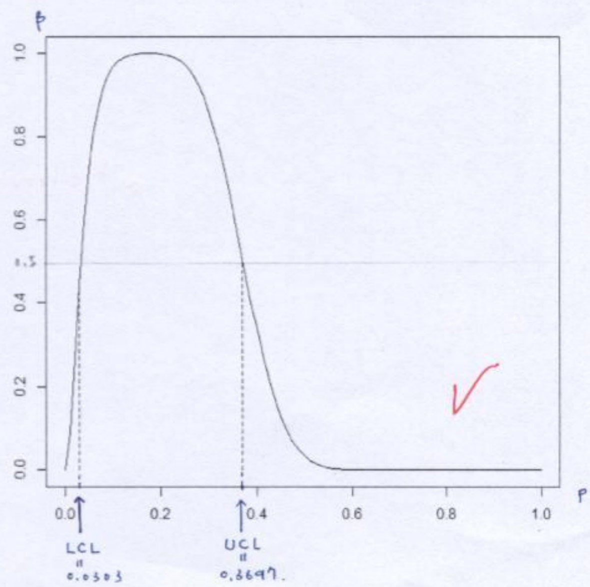
```
> beta=function(p)                          #a function to calculate beta w.r.t p
+ {
+ lfn=log(factorial(50))
+ lp=log(p)
+ l1_p=log(1-p)
+ temp=numeric()
+ for(i in 2:18){                           #the log value of pdf of b(50,p) with x=i
+ lfn_x=log(factorial(50-i))
+ lfx=log(factorial(i))
+ temp[i-1]=lfn-lfn_x-lfx+i*lp+(n-i)*l1_p
+ }
+ temp=exp(temp)
+ return(sum(temp))
+ }
```

why do you use "log"?

And then draw the OC curve for n=50 by using computer

```
> x=seq(0,1,by=0.01)
> y=numeric()
> for(i in 1:101){
+ y[i]=beta(x[i])
+ }
> plot(x,y,xlim=c(0,1),ylim=c(0,1),type="l")
> lines(rep(0.0303,2),c(0,beta(0.0303)),lty=2)
> lines(rep(0.3697,2),c(0,beta(0.3697)),lty=2)
```





In hand-writing one, we draw the line only by 4 points, (0,0), (1,0), (0.0303, $\frac{1}{2}$), (0.3697, $\frac{1}{2}$). But in the other one, it's drawn by 101 points. So the OC curve drawn by computer is more precise.

Q3: +3/2

1) **<Code>:**

```
> library(qcc)
Loading required package: MASS
Package 'qcc', version 2.2
Type 'citation("qcc")' for citing this R package in publications.
> circuits=c(20,20,23,25,12)
> circuits.p=qcc(data=circuits,sizes=200,type="p")
> summary(circuits.p)

Call:
qcc(data = circuits, type = "p", sizes = 200)

p chart for circuits

Summary of group statistics:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.060   0.100   0.100   0.100   0.115   0.125

Group sample size:  200
Number of groups:  5
Center of group statistics:  0.1
Standard deviation:  0.3

Control limits:
       LCL        UCL
 0.03636039 0.1636396
```

+1

p-control chart