

R Lab Session : Part 2 (Solutions)

To see a review of how to start R, look at the beginning of [Lab1](http://www-stat.stanford.edu/~epurdom/RLab.htm)
 Lab1 <http://www-stat.stanford.edu/~epurdom/RLab.htm>

Probability Calculations

The following examples demonstrate how to calculate the value of the cumulative distribution function at (or the probability to the left of) a given number.

- **Normal(0,1) Distribution :**

```
> x <- c(-2, -1, 0, 1, 2)
> x
[1] -2 -1  0  1  2
> pnorm(x)
[1] 0.02275013 0.15865525 0.50000000 0.84134475 0.97724987
```

- **Binomial(n, p) Distribution :**

```
> x <- c(0, 1, 2, 5, 8, 10, 15, 20)
> pbinom(x, size=20, prob=.2)
[1] 0.01152922 0.06917529 0.20608472 0.80420779 0.99001821 0.99943659 0.99999999
[8] 1.00000000
```

- **Poisson(λ) Distribution :**

```
> x <- c(0, 1, 2, 5, 8, 10, 15, 20)
> ppois(x, 6)
[1] 0.002478752 0.017351265 0.061968804 0.445679641 0.847237494 0.957379076
[7] 0.999490902 0.999998545
```

Exercise : Calculate the following probabilities :

1. Probability that a normal random variable with mean 22 and variance 25
 - (i) lies between 16.2 and 27.5
`pnorm(27.5, 22, sd=5) - pnorm(16.2, 22, sd=5)`
`[1] 0.7413095`
 - (ii) is greater than 29 `1 - pnorm(29, 22, sd=5)`
`[1] 0.08075666`
 - (iii) is less than 17 `pnorm(17, 22, sd=5)`
`[1] 0.1586553`
 - (iv) is less than 15 or greater than 25 `pnorm(15, 22, sd=5) + 1 - pnorm(25, 22, sd=5)`
`[1] 0.3550098`

2. Probability that in 60 tosses of a fair coin the head comes up

- (i)

20,25 or 30 times

```
sum(dbinom(c(20,25,30),60,prob=0.5))
[1] 0.1512435
```

(ii)

less than 20 times

```
pbinom(19,60,prob=0.5)
[1] 0.0031088
```

(iii)

```
between 20 and 30 times pbinom(30,60,prob=0.5)-pbinom(20,60,prob=0.5)
[1] 0.5445444
```

3.

A random variable X has Poisson distribution with mean 7. Find the probability that

(i)

X is less than 5 **less or equal is:**

```
> ppois(5,7)
[1]0.3007083
less than is
> ppois(4,7)
[1]0.1729916
```

(ii)

X is greater than 10 (strictly)

```
> 1-ppois(10,7) [1] 0.0985208
```

(iii)

```
X is between 4 and 16 > ppois(16,7)-ppois(3,7) [1] 0.9172764
```

Quantiles

The following examples show how to common the quantiles of some common distributions for a given probability (or a number between 0 and 1).

- **Normal(0,1) Distribution :**

```
> y <- c(.01,.05,.1,.2,.5,.8,.95,.99)
> qnorm(y,mean=0,sd=1)
[1] -2.3263479 -1.6448536 -1.2815516 -0.8416212 0.0000000 0.8416212 1.6448536
[8] 2.3263479
```

- **Binomial(n, p) Distribution :**

```
> y <- c(.01,.05,.1,.2,.5,.8,.95,.99)
> qbinom(y,size=30,prob=.2)
[1] 1 3 3 4 6 8 10 11
```

- **Poisson(λ) Distribution :**

```
> y <- c(.01,.05,.1,.2,.5,.8,.95,.99)
> qpois(y,6)
[1] 1 2 3 4 6 8 10 12
```

Random Variable generation

The following examples illustrate how to generate random samples from some of the well-known probability distributions.

- **Normal(μ, σ^2) Distribution :**

The first sample is from $N(0, 1)$ distribution and the next one from $N(5, 1)$ distribution.

```
> z <- rnorm(10)
> z
[1] -0.90361592 -1.96522764 -1.35107949 -0.10846423 0.29756634 1.40831606
[7] -0.07844737 1.40575257 -0.97511415 -0.33418299
```

If you would like to see how the distribution of the sample points looks like

```
> w <- rnorm(1000, mean=5, sd=1)
> hist(w)
```

- **Binomial(n, p) Distribution :**

```
> k <- rbinom(20, size=5, prob=.2)
> k
[1] 1 2 0 1 0 0 0 2 0 1 0 0 0 0 2 4 1 1 1
```

- **Poisson(λ) Distribution :**

```
> x <- rpois(20, 6)
> x
[1] 2 8 7 5 5 5 3 8 5 5 1 8 5 5 5 4 10 7 3 4
```

Exercise (Advanced) : Generate 500 samples from Student's t distribution with 5 degrees of freedom and plot the histogram. (Note: t distribution is going to be covered in class). The corresponding function is **rt** .
`hist(rt(500, 5), 40)`

Density Plots

- **Plotting the probability density function (pdf) of a Normal distribution :**

```
> x11()
> x <- seq(-4.5, 4.5, .1)
> normdensity <- dnorm(x, mean=0, sd=1)
> plot(x, normdensity, type="l")
```

- **Plotting the probability mass function (pmf) of a Binomial distribution :**

```
> par(mfrow=c(2,1))
> k <- c(1:30)
> plot(k, dbinom(k, size=30, prob=.15), type="h")
> plot(k, dbinom(k, size=30, prob=.4), type="h")
> par(mfrow=c(1,1))
```

- **Discrete Probabilities** For a discrete random variable, you can use the probability mass to find $P(X = k)$

```
> dbinom(3, size=10, prob=0.5)
[1] 0.1171875
```

**** Note the distinction between the continuous (Normal) and the discrete (Binomial) distributions.**

Exercise : Plot the probability mass functions for the Poisson distribution with mean 4.5 and 12 respectively. Do you see any similarity of these plots to any of the plots above? If so, can you guess why ?

Exercise : Recreate the probabilities that Professor Holmes did in class (Bin(5,.4)) [You can do it in 1 command!] How would you get the expected counts?

Q-Q_plot

R has two different functions that can be used for generating a Q-Q plot. Use the function **qqnorm** for plotting sample quantiles against theoretical (population) quantiles of standard normal random variable.

Example :

```
> stdnormsamp <- rnorm(100, mean=0, sd=1)
> normsamp <- rnorm(100, mean=5, sd=1)
> binsamp <- rbinom(100, size=20, prob=.25)
> poissamp <- rpois(100, 5)

> par(mfrow=c(2,2))

> qqnorm(stdnormsamp, main="Normal Q-Q plot : N(0,1) samples")
> qqline(stdnormsamp, col=2)
> qqnorm(normsamp, main="Normal Q-Q plot : N(5,1) samples")
> qqline(normsamp, col=2)
> qqnorm(binsamp, main="Normal Q-Q plot : Bin(20,.25) samples")
> qqline(binsamp, col=2)
> qqnorm(poissamp, main="Normal Q-Q plot : Poisson(5) samples")
> qqline(poissamp, col=2)
```

Note: Systematic departure of points from the Q-Q line (the red straight line in the plots) would indicate some type of departure from normality for the sample points.

Use of function **qqplot** for plotting sample quantiles for one sample against the sample quantiles of another sample

Example :

```
> par(mfrow=c(2,1))

> qqplot(stdnormsamp, normsamp, xlab = "Sample quantiles : N(0,1) samples",
+ ylab = "Sample quantiles : N(5,1) samples")
> qqplot(stdnormsamp, binsamp, xlab = "Sample quantiles : N(0,1) samples",
+ ylab = "Sample quantiles : Bin(20,.25) samples")
```

Exercise : Generate 100 samples from Student's t distribution with 4 degrees of freedom and generate the qqplot for this sample.

qqnorm(rt(100, df=4)) Generate another sample of same size, but now from a t distribution with 30 degrees of freedom and generate the q-q plot. Do you see any difference ?

qqnorm(rt(100, df=30))

It should be evident to you that the t distribution is very far from normal, and the 30 degrees of freedom t is indistinguishable from Normal.

Susan Holmes 2004-10-31