Tutorial 4: Statistical Theory; Why is the Gaussian Distribution so popular? – Answers Sheet

Rob Nicholls – nicholls@mrc-lmb.cam.ac.uk

MRC LMB Statistics Course 2014

Task 1

1.

```
pnorm(9,10,sqrt(4))
```

Answer: 0.309 (3sf)

2.

$$
\begin{aligned}
P(|X-8| \leq 1) &= P(-1 \leq X-8 \leq 1) \\
&= P(7 \leq X \leq 9) \\
&= P(X \leq 9) - P(X \leq 7)
\end{aligned}
$$

```
pnorm(9,10,sqrt(4)) - pnorm(7,10,sqrt(4))
```

Answer: 0.242 (3sf)

Task 2

$$
\begin{aligned}
Cov(X,Y) &= E\left((X-\mu_X)(Y-\mu_Y)\right) \\
&= E\left(XY - X\mu_Y - \mu_X Y + \mu_X\mu_Y\right) \\
&= E\left(XY\right) - E\left(X\mu_Y\right) - E\left(\mu_X Y\right) + E\left(\mu_X\mu_Y\right) \\
&= E\left(XY\right) - E\left(X\right)\mu_Y - \mu_X E\left(Y\right) + \mu_X\mu_Y \\
&= E\left(XY\right) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\
&= E\left(XY\right) - \mu_X\mu_Y
\end{aligned}
$$

Task 3

1.

```
x=rbinom(10,100,0.4)
```

2.

```
plot(density(x))
```

3.

If $X \sim \text{Bin}(n, p)$ then $E(X) = np$ and $Var(X) = np(1-p)$.

If $X \sim \text{Bin}(100, 0.4)$ then $E(X) = 40$ and $Var(X) = 24$.

4.

```
y=0:100
plot(dnorm(y,40,sqrt(24))~y)
```

5.

```
lines(density(x))
```

The density functions are in roughly the same place, but they do not overlay particularly well.

6.

```
lines(density(rbinom(100,100,0.4)))
```

With more observations, the functions better overlap. Consequently, we can see that the Normal distribution appears to approximate the Binomial distribution reasonably well. However, there are still some differences between the two curves, implying that the approximation may not be perfect.

7.

```
for(i in 1:100){lines(density(rbinom(100,100,0.4)))}
```

It is clear that, even using 100 observations per trial, there is a relatively large degree of uncertainty associated with the density functions corresponding to the simulated Binomial samples. Furthermore, we can see that the density function corresponding to the Normal approximation seems to lie well within the range of density functions corresponding to the Binomial samples. Consequently, we conclude that the Normal approximation should be considered a very good approximation to the Binomial (for $0 << p << 1$), given that we have observed it to 'perform better' than many true random Binomial samples containing (even as many as) 100 observations.

8.

```
pnorm(35,40,sqrt(24))
```

Answer: 0.154 (3sf)

9.

```
pbinom(35,100,0.4)
```

Answer: 0.179 (3sf)
Whilst not perfect, the approximation is reasonably close to the true value.

10.

```
x=c()
for(i in 1:100){
x=c(x,sum(rbinom(100,100,0.4)<=35)/100)
}

plot(density(x))
mean(x)
sd(x)
```

The true value from the Binomial distribution (0.179) lies near the centre of this distribution. The value from the Normal approximation (0.154) lies well within this distribution. Whilst not right in the centre, this value (0.154) lies well within two standard deviations of the mean (i.e. the approximate 95% confidence interval). Consequently, we conclude that the Normal approximation can be considered a very good approximation to the Binomial, given that it could have arisen by chance from a sample of many (100) random variates from the appropriate Binomial distribution.

Task 5

To prove that $Y \sim N(\alpha\mu+\beta, \alpha^2\sigma^2)$, we have to show that Y is Normally-distributed with mean $\hat{\mu} = \alpha\mu + \beta$ and variance $\hat{\sigma}^2 = \alpha^2\sigma^2$. We can prove this if we can show that the moment generating function of Y is that of a Normal random variable, i.e.:

$$m_Y(t) = e^{t\hat{\mu}+\frac{1}{2}t^2\hat{\sigma}^2}$$

Since $Y = \alpha X + \beta$ we have:

$$\begin{aligned}
m_Y(t) &= m_{\alpha X+\beta}(t) \\
&= E(e^{(\alpha X+\beta)t}) \\
&= E(e^{\alpha Xt}e^{\beta t}) \\
&= e^{\beta t}E(e^{\alpha Xt}) \\
&= e^{\beta t}m_X(\alpha t) \\
&= e^{\beta t}e^{\alpha t\mu+\frac{1}{2}(\alpha t)^2\sigma^2} \\
&= e^{(\alpha\mu+\beta)t+\frac{1}{2}t^2\alpha^2\sigma^2} \\
&= e^{\hat{\mu}t+\frac{1}{2}t^2\hat{\sigma}^2}
\end{aligned}$$

which is the moment generating function of a Normally-distributed random variable with mean $\hat{\mu}$ and variance $\hat{\sigma}^2$, proving that $Y \sim N(\hat{\mu}, \hat{\sigma}^2)$, as required.