

# MATH 308 Assignment 8:

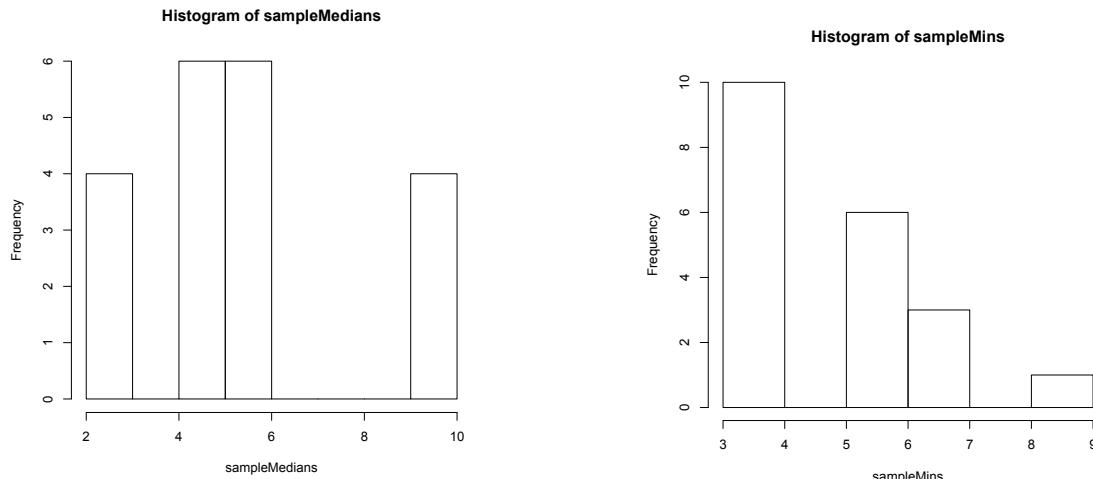
## Exercises 4.4

Nakul Joshi

March 10, 2014

1

2



Population median: 5.5

Mean of sample medians: 5.7

Mean of minimums: 4.8

6

From the CLT, the sample mean is approximately distributed normally with  $\mu_{\bar{X}} = \mu = 48$  and  $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = 9^2/20 = 2.7$ . Thus the probability of the sample mean being greater than 51 is `1-pnorm(51,mean=48,sd=sqrt(2.7))` which evaluates to  $\approx 3\%$ .

## 9

The mean  $\mu$  of the distribution is  $\int_2^6 f(x) dx = 4$ . The variance  $\sigma^2$  is  $\int_2^6 (x - \mu)^2 f(x) dx = 2.4$ . By CLT, the distribution of the sample mean  $\bar{X}$  is approximated by  $\mathcal{N}(\mu, \frac{\sigma^2}{n}) = \mathcal{N}(4, 3/305)$ . Thus,  $P(\bar{X} \geq 4.2)$  is  $1 - \text{pnorm}(4.2, \text{mean}=4, \text{sd}=\sqrt{2.4})$ , which evaluates to  $\approx 2\%$ .

## 10

By CLT, the number of people with degrees is approximately normally distributed with  $\mu = np$  and  $\sigma^2 = np(1-p)$ . Thus, the probability of between 220 and 230 people in the sample having a degree is:

```
n=800
p=0.286
m=n*p
sd=sqrt(n*p*(1-p))
pn=
pnorm(230+0.5 ,m ,sd)
-pnorm(220-0.5 ,m ,sd)
```

Which evaluates to 0.3194833. The exact probability is given by

```
pb=pbinom(230 ,n ,p)-pbinom(220 ,n ,p)
```

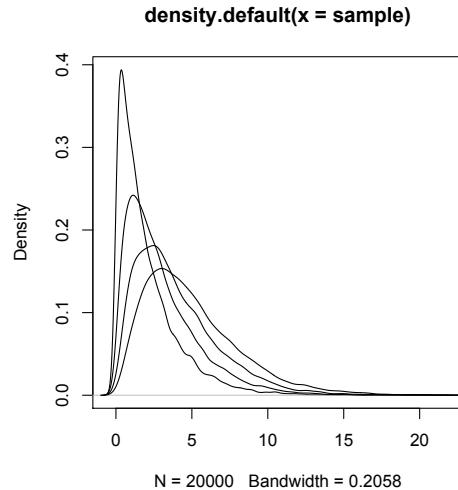
which evaluates to 0.2959644. The error in the approximation is  $(pn - pb)/pb$  which evaluates to  $\approx 8\%$ .

## 15

By definition,  $W \sim \chi_n^2$ . The sampling distributions of  $W_n$  for  $n \in \{2, 3, 4, 5\}$  are shown in the accompanying figure. The means and variances are tabulated as:

$n$	$\mu$	$\sigma^2$
2	2.02	4.07
3	2.99	5.82
4	3.98	7.80
5	4.99	10.01

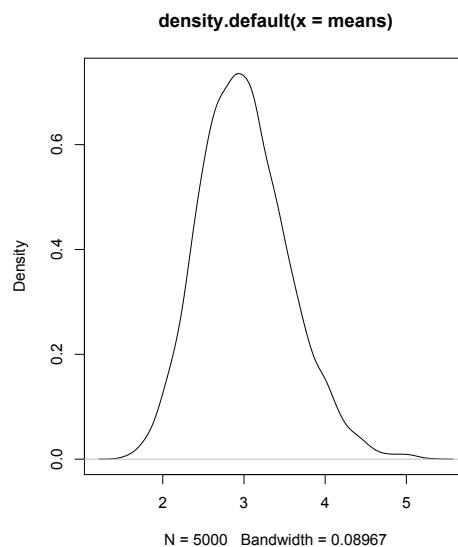
From the table, we can see that  $\mu = n$  and  $\sigma^2 = 2n$ .



## 18

a)

The sampling distribution is shown below:



b)

25

The mean from sample is 3.00 with standard error 0.55. From CLT, we expect  $\mu_{\bar{X}} = \mu = (1/3)^{-1} = 3$  and  $\sigma_{\bar{X}} = \sigma/\sqrt{n} = (1/3)^{-1}/\sqrt{30} \approx 0.55$ , giving a difference of < 1%.

c)

We get  $P = 0.83$ .

d)

We get  $P = 0.82$ , which is only  $\approx 1\%$  off.

20

a) Distribution of Sample Minimum

$$\begin{aligned} P(X_{\min} \geq x) &= \prod_{i=1}^n P(X_i \geq x) \\ \implies 1 - F_{\min}(x) &= (1 - F(x))^n \\ \implies -f_{\min}(x) &= n(1 - F(x))^{n-1}(-f(x)) \\ \implies f_{\min}(x) &= n(1 - F(x))^{n-1}f(x) \quad \square \end{aligned}$$

b) Distribution of Sample Maximum

$$\begin{aligned} P(X_{\max} \leq x) &= \prod_{i=1}^n P(X_i \leq x) \\ \implies F_{\max}(x) &= F^n(x) \\ \implies f_{\max}(x) &= nF^{n-1}(x)f(x) \quad \square \end{aligned}$$

22

a)

$F(x) = \int_1^x \frac{2}{x^2} dx = 2 - 2/x$ , so from previous result  $f_{\max} = 2 \left(2 - \frac{2}{x}\right)^{2-1} \frac{2}{x^2} = 2 \left(2 - \frac{2}{x}\right) \frac{2}{x^2}$ .

b)

$$E_{X_{\max}} = \int_1^2 x f_{\max}(x) dx \approx 1.55.$$

$$\begin{aligned} f_{X_1+X_2}(x) &= P(X_1 + X_2 = x) \\ &= \sum_{i=0}^x \left( \frac{\lambda_1^i}{i!} e^{-\lambda_1} \frac{\lambda_2^{x-i}}{(x-i)!} e^{-\lambda_2} \right) \\ &= e^{-(\lambda_1+\lambda_2)} \sum_{i=0}^x \frac{\lambda_1^i \lambda_2^{x-i}}{i!(x-i)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{x!} \sum_{i=0}^x \binom{x}{i} \lambda_1^i \lambda_2^{x-i} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{x!} (\lambda_1 + \lambda_2)^x \\ &= f_Z(x), \text{ where } Z \sim \mathcal{P}(\lambda_1 + \lambda_2) \end{aligned}$$

$$\begin{aligned} &\implies X \sim \mathcal{P} \left( \sum_{i=1}^{10} \lambda_i \right) \\ &\implies X \sim \mathcal{P}(30) \\ &\implies f_X(x) = \frac{30^x}{x!} e^{30} \end{aligned}$$

27

a)

The distribution seems quite close to normal.

b)

Theoretical mean  $\mu_{\bar{X}} = \mu = 10.17$ . Actual mean is 10.02. Difference is  $\approx 1.5\%$ .

c)

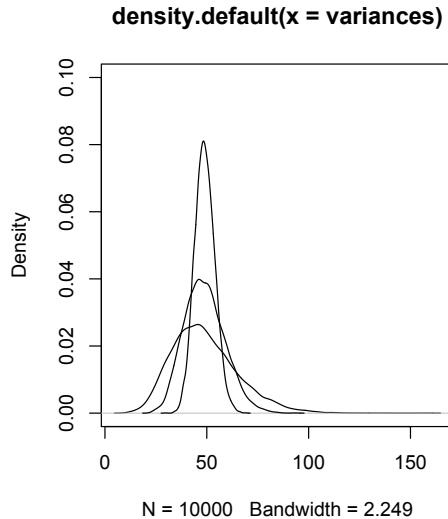
Theoretical s.e.  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \approx 4.55$ . Actual s.e. 4.56. Difference is < 1%.

d)

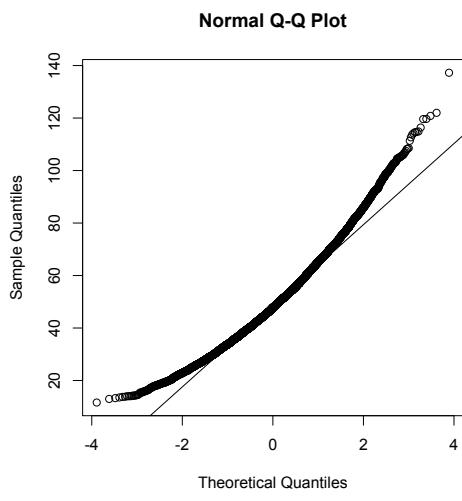
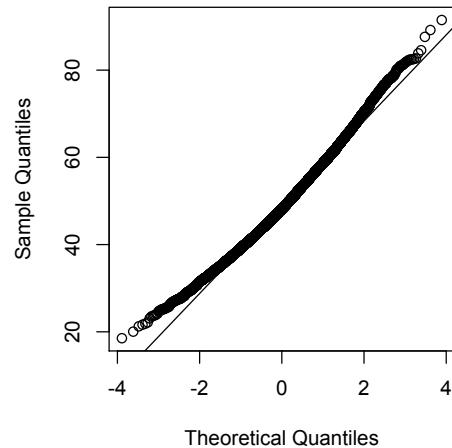
The differences reduce with increasing  $n$ .

## 28

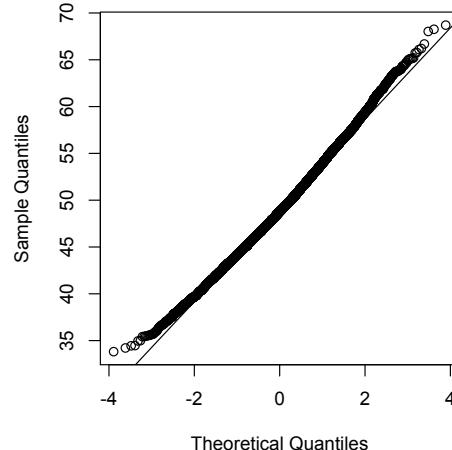
It can be seen from the plots below that the distribution of sample variances is approximately normal, and the fit to normal improves with sample size.



**Normal Q-Q Plot**



**Normal Q-Q Plot**



# Math 308 Assignment 9

## Moments of the Standard Normal

Nakul Joshi

March 6, 2014

### 1 Moment Generating Function

From the definition of the m.g.f.,

$$\begin{aligned} M_Z(t) &= \mathbb{E}(e^{tZ}) \\ &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} e^{t^2/2} dx \\ &= e^{t^2/2} \int_{-\infty}^{\infty} f_X(x) dx \end{aligned}$$

where  $X \sim \mathcal{N}(t, 0)$

$$= e^{t^2/2}$$

### 2 Moments

Using the series expansion of the exponential

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

We observe that

$$\begin{aligned} M_Z(t) &= \sum_{i=0}^{\infty} \frac{(t^2/2)^i}{i!} \\ &= 1 + \frac{t^2}{2} + \frac{t^4}{2^2 2!} + \frac{t^6}{2^3 3!} + \frac{t^8}{2^4 4!} \dots \\ M'_Z(t) &= \frac{2t}{2} + \frac{4t^3}{2^2 2!} + \frac{6t^5}{2^3 3!} + \frac{8t^7}{2^4 4!} \dots \\ M''_Z(t) &= 1 + \frac{4 \times 3t^2}{2^2 2!} + \frac{6 \times 5t^4}{2^3 3!} + \frac{8 \times 7}{2^4 4!} \dots \end{aligned}$$

This tells us that, for odd values of  $n$ ,  $M_Z^{(n)}(t)$  will be an expression of the form  $t \times (\text{a polynomial})$ . Thus, for these values,  $M_Z^{(n)}(0)$  is 0.

For the even values of  $n$ , we can see that  $M_Z^{(n)}(t)$  is simply the  $n$ th derivative of the  $\frac{n}{2}$ th term in the series expansion of  $M_Z(t)$ , plus an expression of the form  $t \times (\text{a polynomial})$ . The latter expression always evaluates to 0 at  $t = 0$ , so we can drop it. To get the  $n$ th derivative of the  $i$ th term:

$$\begin{aligned} T_i &= \frac{t^{2i}}{2^i i!} \\ \frac{d^n}{dt^n} T_i &= \frac{(2i)(2i-1)(2i-2)\dots(2i-n+1)}{2^i i!} t^{2i-n} \\ &= \frac{(2i)!}{(2i-n)!} \frac{1}{2^i i!} t^{2i-n} \end{aligned}$$

Setting  $i$  to  $n/2$ ,

$$\begin{aligned} \mathbb{E}(Z^n) &= \frac{n!}{0!} \frac{1}{2^{(n/2)} (n/2)!} t^0 \\ &= \frac{n!}{(n/2)!} \frac{1}{2^{n/2}} \end{aligned}$$

We can then compute and tabulate the moments:

$n$	0	1	2	3	4	5	6
$\mathbb{E}(Z^n)$	1	0	1	0	3	0	15

# Math 308 Assignment 10

## Mean and Variance of the Chi-Squared Distribution

Nakul Joshi

March 6, 2014

### 1 Moment Generating Function

From the definition of the m.g.f.,

$$M_X(t) = \mathbb{E}(e^{tX})$$

$$= \int_0^\infty e^{tx} \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} dx$$

But,  $e^{-\frac{x}{2}} e^{tx} = e^{-x(t-\frac{1}{2})}$ .

Substituting  $u = x(1/2 - t)$ , we get  $du = (1/2 - t)dx$ ,  
and  $u|_{x=0} = 0$ ,  $u|_{x=\infty} = \infty$ . So, we get:

$$\begin{aligned} M_Z(t) &= \int_0^\infty \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} \frac{u^{\frac{k}{2}-1}}{(1/2 - t)^{\frac{k}{2}-1}} \frac{e^{-u}}{1/2 - t} du \\ &= \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} \frac{1}{(1/2 - t)^{\frac{k}{2}}} \int_0^\infty u^{\frac{k}{2}-1} e^{-u} du \\ &= \frac{1}{2^{k/2}\cancel{\Gamma(\frac{k}{2})}} \frac{1}{(1/2 - t)^{\frac{k}{2}}} \cancel{\Gamma(\frac{k}{2})} \\ &= (1 - 2t)^{-\frac{k}{2}} \end{aligned}$$

### 2 Mean and Variance

From the above m.g.f., we can calculate the derivatives

$$M'_X(t) = k(1 - 2t)^{-\frac{k}{2}-1}, \text{ and}$$

$$M''_X(t) = k(k + 2)(1 - 2t)^{-\frac{k}{2}-2}.$$

This gives

$$\mu = \mathbb{E}(X) = M'_X(0) = k$$

We can also calculate

$$\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$= M''_X(0) - k^2$$

$$= k(k + 2) - k^2$$

$$= 2k$$

# MATH 308 Assignment 11:

## Baseball World Series Hypothesis Testing

Nakul Joshi

March 6, 2014

### 1 Expected Values under the Null

### 2 Hypothesis Testing

To win the series after playing  $n$  games, the winning team must win 3 of the first  $n-1$  games, and then win the last game. Thus, for each team, the probability of winning after playing  $n$  games is:

$$P(n = k) = \binom{k-1}{3} \left(\frac{1}{2}\right)^{k-1} \left(\frac{1}{2}\right)$$

However, the probability of  $n$  games being played is the probability of *either* team winning after  $n$  games. Thus, the p.m.f. of the number of games played,  $N$ , is twice the function above:

$$P(N = k) = \binom{k-1}{3} \left(\frac{1}{2}\right)^{k-1}$$

Thus, the expected number of series  $N$  won after  $k$  games is  $P(N_k) \times (\text{number of games played})$ . We can then tabulate the expected and observed values:

$k$	4	5	6	7
$\mathbb{E}(N_k)$	6.25	12.5	15.625	15.625
$\mathbb{O}(N_k)$	8	8	10	24

The test statistic  $t$  is given by  $\sum_{k=4}^7 \frac{(\mathbb{O}(N_k) - \mathbb{E}(N_k))^2}{\mathbb{E}(N_k)} = 9.76$ . Because  $k$  can take on only 4 values, we know that  $t \sim \chi_3^2$ . Hence, our  $p$  value is given by

$$p = \int_{9.76}^{\infty} \frac{e^{-x/2} \sqrt{x}}{\sqrt{2\pi}} dx \approx 0.02$$

Thus, we can reject the null at the 5% significance level.