STA 250/MTH 342 Intro to Mathematical Statistics
Lab Session 8 / March 23, 2015 / Handout

In this lab session we study the computation of confidence intervals.

See: `https://stat.duke.edu/courses/Spring15/sta250/labs/` for links to source code and data. Submit lab solutions via email to: `sta250@stat.duke.edu`. Any plots should be included in postscript form as attachments. The email subject must be "STA250 ..." with "..." replaced by your name.

1: Confidence interval for the mean of normal distribution. Suppose we have 12 numbers

`X = {4.2, 4.6, 2, 0.9, 4.2, -3.4, 4.3, 1.5, 1.4, -0.58, 2, 8.1},`

which are drawn independently from the normal distribution $\mathsf{No}(\mu, \sigma^2 = 10)$ with unknown μ. We now find the confidence interval for μ with confidence coefficient 0.99.

First, the MLE for μ is $\hat{\mu} = \bar{X}_n$, which has a sampling distribution $\mathsf{No}(\mu, \sigma^2/n)$. One may use the transformation $Z := \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}$ to find the confidence interval $[\bar{X}_n - \frac{\Phi^{-1}(0.995)\sigma}{\sqrt{n}}, \bar{X}_n + \frac{\Phi^{-1}(0.995)\sigma}{\sqrt{n}}]$.

```
> X <- c(4.2, 4.6, 2, 0.9, 4.2, -3.4, 4.3, 1.5, 1.4, -0.58, 2, 8.1);
> ci.lower <- mean(X) - qnorm(0.995)*sqrt(10)/sqrt(12);
> ci.lower;
[1] 0.08360031
> ci.upper <- mean(X) + qnorm(0.995)*sqrt(10)/sqrt(12);
> ci.upper;
[1] 4.7864
```

Alternatively, one may build the confidence interval directly from $\mathsf{No}(\mu, \sigma^2/n)$.

```
> ci.lower <- qnorm(0.005, mean = mean(X), sd = sqrt(10)/sqrt(12));
> ci.lower;
[1] 0.08360031
> ci.upper <- qnorm(0.995, mean = mean(X), sd = sqrt(10)/sqrt(12));
> ci.upper;
[1] 4.7864
```

TASK 1 Suppose the variance is unknown. Make a confidence interval with the same confidence coefficient. Note, that now one may use $Y := \sqrt{12}(\bar{X}_{12} - \mu)/s$, which has the t_{11} distribution, where $s^2 := \frac{1}{11}\sum_{i=1}^{12}(X_i - \bar{X}_{12})^2$ is the sample variance. One may get the quantile by

```
> qt(0.025, df = 11);
[1] -2.200985
> qt(0.975, df = 11);
[1] 2.200985
```

2: Confidence interval for the variance of normal distribution. Suppose the above 12 numbers are drawn independently from $\mathsf{No}(\mu, \sigma^2)$ where both μ and σ^2 are unknown. We now find a confidence interval for σ^2 with confidence coefficient 0.95.

We have derived that $Z := \frac{11s^2}{\sigma^2}$ has a χ^2_{11} distribution. Let c_1 and c_2 be any two positive numbers such that

$$\Pr\left[c_1 < \frac{\sum_{i=1}^{12}(X_i - \bar{X}_{12})^2}{\sigma^2} < c_2\right] = 0.95,$$

then $[\frac{11s^2}{c_2}, \frac{11s^2}{c_1}]$ will be the confidence interval we want (note the order). We will compute the "symmetric" interval by choosing c_1 and c_2 so to be the 0.025 and 0.975 quantiles respectively.

```
> ci.lower <- sum((X - mean(X))^2) / qchisq(0.975, df = 11);
> ci.lower;
[1] 4.302258
> ci.upper <- sum((X - mean(X))^2) / qchisq(0.025, df = 11);
> ci.upper;
[1] 24.71486
```

TASK 2 Suppose that $\mu = 1.50$ is known, for a sample of size $n = 12$ from the $\mathsf{No}(\mu, \sigma^2)$ distribution. Now find a symmetric 95% confidence interval for σ^2.

3: Computing quantiles empirically. We see from the above example that in many cases, finding the confidence interval is straight forward, so long as one can find the quantile. Recall that in last lab session, we mentioned the following inverse gamma distribution, $\mathsf{InvGamma}(\alpha, \beta)$ where the parameters α and β should both be positive. The pdf is

$$f_{\mathsf{InvGamma}(\alpha,\beta)}(x) \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x}, & \text{for } x > 0, \\ 0, & \text{for } x \le 0. \end{cases}$$

Recall the definition of $\mathsf{Ga}(\alpha, \beta)$, where $\alpha, \beta > 0$. The pdf is

$$f_{\mathsf{Gamma}(\alpha,\beta)}(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & \text{for } x > 0, \\ 0, & \text{for } x \le 0. \end{cases}$$

One has

$$X \sim \mathsf{Ga}(\alpha, \beta) \Leftrightarrow \frac{1}{X} \sim \mathsf{InvGamma}(\alpha, \beta).$$

Now we find the 0.975 and 0.025 quantiles for the inverse gamma distribution $\mathsf{InvGamma}(\alpha = 1.5, \beta = 2.3)$ empirically.

```
> z <- 1 / rgamma(n = 10000, shape = 1.5, rate = 2.3);
> quantile(z, probs = c(0.025, 0.975));
```

```
      2.5%      97.5%
 0.4834423 21.2013814
> quantile(1 / rgamma(n = 1000000, shape = 1.5, rate = 2.3), c(0.025, 0.975));
      2.5%      97.5%
 0.4916671 21.3636726
```

For verifying the accuracy, we numerically solve the equation for x,

$$\int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} t^{-(\alpha+1)} e^{-\beta/t}\, \mathrm{d}t = 0.025.$$

```
> install.packages("nleqslv");
> library("nleqslv");
> alpha <- 1.5;
> beta  <- 2.3;
> f <- function(t)beta ^ alpha / gamma(alpha) * t^( - alpha - 1) * exp( - beta / t);
> g <- function(x)integrate(f, 0, x)$value - 0.025;
> nleqslv(x = 0.4, fn = g)$x;
[1] 0.4920626
> h <- function(x)integrate(f, 0, x)$value - 0.975;
> nleqslv(x = 0.4, fn = h)$x;
[1] 21.31649
```

We see that our empirical approximation is really precise.

TASK 3 Recall that if we have a sample $X_1, \cdots, X_n$ drawn independently from $\mathsf{No}(\mu, \sigma^2)$, then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{s} \sim t_{n-1}.$$

Please use this to generate 10,000 independent samples from t_{10}, and find empirically the 0.05 and 0.95 quantiles. Compare them with the true values.

~~END~~