Estimation of Parameters and Statistical Sampling

November 6, 2010

1 Introduction

Here we consider two types of statistical sampling problems, one is just for pedagogical purposes, the other is directly applicable to real problems. These two problems are:

Problem 1 (pedagogical). Suppose that X is a random variable for which we know the variance σ^2, but do not know the mean μ. One way to estimate μ would be to take samples of X, and then average. That is, suppose that $X_1, ..., X_k$ are independent random variables with the same distribution as X; then, we let

$$\hat{\mu} \ = \ \frac{X_1 + \cdots + X_k}{k}$$

be an estimator for μ. Note that $\hat{\mu}$ is a random variable, and for large values of k it will have approximately a normal distribution with mean μ (by the Central Limit Theorem).

The sort of thing we would like to compute is a 95% confidence interval for μ, which is an interval $(\hat{\mu}-\delta, \hat{\mu}+\delta)$ such that 95% of the time (remember, $\hat{\mu}$ is a random variable), μ lies in this interval.

The reason that this problem is only pedagogical is that in real world problems we are unlikely to encounter situations where we know σ, but not μ.

Problem 2 (real). This is the exact same problem, except that here we know neither μ nor σ; *in addition*, we will assume that X is normal (a standard assumption for many statistical sampling problems). This problem is vastly more difficult to analyze theoretically; however, we are in luck that it was worked out long ago. There is actually a nice little bit of history surrounding this that we will discuss below.

Basically, as before, we suppose that $X_1, ..., X_k$ are independent and have the same distribution as $X = N(\mu, \sigma^2)$, and we consider

$$\hat{\mu} \ = \ \frac{X_1 + \cdots + X_k}{k}$$

and

$$\hat{\sigma}^2 \ = \ \frac{1}{k-1} \sum_{i=1}^{k} (X_i - \overline{X})^2.$$

The problem here is to determine δ such that $(\hat{\mu} - \delta, \hat{\mu} + \delta)$ is a 95% confidence interval for μ; and, we would furthermore like a 95% confidence interval for σ^2 (or just σ).

As with Problem 1, for large values of k it will turn out that $\hat{\mu}$ and $\hat{\sigma}^2$ are approximately normal; however, we would like to be able to say something for when k is small. In a later section we will do this.

2 Problem 1

We know that $\hat{\mu}$ is a maximum likelihood estimator for μ, and that for large k we have that $\hat{\mu}$ is approximately normal, by the central limit theorem. How and why is this the case? Well, from the central limit theorem, we know that for large k,

$$\frac{X_1 + \cdots + X_k - k\mu}{\sigma\sqrt{k}} \ \sim \ N(0,1).$$

What does this mean? It means that for any given real number c, we have that

$$\lim_{k\to\infty} P\left(\frac{X_1 + \cdots + X_k - k\mu}{\sigma\sqrt{k}} \ < \ c\right) \ = \ P(N(0,1) < c) \ = \ \Phi(c).$$

Now, we have that

$$P(\hat{\mu} < c) \quad = \quad P\left(\frac{X_1 + \cdots + X_k}{k} < c\right)$$

$$\begin{aligned}
&= P\left(\frac{X_1 + \cdots + X_k - k\mu}{k} < c - \mu\right) \\
&= P\left(\frac{X_1 + \cdots + X_k - k\mu}{\sigma\sqrt{k}} < \sigma^{-1}(c - \mu)\sqrt{k}\right) \\
&\sim P\left(N(0,1) < \sigma^{-1}(c - \mu)\sqrt{k}\right) \\
&= \Phi(\sigma^{-1}(c - \mu)\sqrt{k}).
\end{aligned}$$

Now, for a 95% confidence interval, we need to compute δ so that

$$(\hat{\mu} - \delta, \hat{\mu} + \delta) \text{ contains } \mu$$

occurs with 95% probability. [1] That is, we seek δ so that

$$\hat{\mu} \in (\mu - \delta, \mu + \delta)$$

with 95% probability. That is, we seek δ to so that

$$\begin{aligned}
0.95 &= \Phi(\sigma^{-1}\delta\sqrt{k}) - \Phi(-\sigma^{-1}\delta\sqrt{k}) \\
&= 2\Phi(\sigma^{-1}\delta\sqrt{k}) - 1.
\end{aligned}$$

For this last step we have used the fact that for $x > 0$,

$$\Phi(-x) = 1 - \Phi(x).$$

So, we seek δ so that

$$\Phi(\sigma^{-1}\delta\sqrt{k}) = \frac{0.95 + 1}{2} = 0.975.$$

This is easy to do via a table lookup.

[1] The reason we don't say that $\mu \in (\hat{\mu} - \delta, \hat{\mu} + \delta)$ is that it sounds like one is saying that μ is a random variable, when in fact μ is a constant; $\hat{\mu}$ is the random variable.

3 Problem 2

Even if we assume that k is large, we cannot use the idea from the previous section to determine a confidence interval for μ without knowing σ, because our confidence interval formula given above involves σ. Even the Central Limit Theorem is of no use in this case. However, we can try to estimate σ^2 using the estimator $\hat{\sigma}^2$. But then, it is not immediately clear to what degree this affects the size of our confidence interval when k is small, say around 30. In this section we will address these problems.

The theorem we will use to obtain confidence intervals is:

Theorem 1 *Let*

$$t \;=\; \frac{(\overline{X}-\mu)\sqrt{k}}{\hat{\sigma}}.$$

Then, t has a Student-t distribution with $k-1$ degrees of freedom. That is, t has the following pdf:

$$f(x) \;=\; \frac{\Gamma(k/2)}{\Gamma((k-1)/2)\sqrt{\pi(k-1)}}\left(1+\frac{x^2}{k-1}\right)^{-k/2}.$$

And, if we let

$$v \;=\; \frac{(k-1)\hat{\sigma}^2}{\sigma^2},$$

then $v \sim \chi^2_{k-1}$; that is, v has a χ^2 distribution with $k-1$ degrees of freedom.

And now a little bit of history regarding the student-t distribution: It was worked out in the early 1900's by a statistician named William Sealy Gosset, who worked for the beer company *Guinness*. Basically, Gosset developed it as a way to handle the problem of "small sample sizes" that brewers had to work with. Because Gosset's result was a trade secret of the company, which meant he couldn't publish it under his true name, he used the pseudonym "Student t". See the following wikipedia page for more details:

`http://en.wikipedia.org/wiki/William_Sealy_Gosset`

3.1 Student t is approximately $N(0,1)$ for large k

Here, we will show that t approaches $N(0,1)$ in distribution as $k \to \infty$. Basically, we need to see how the ratio of these gamma factors behaves as k tends to infinity. To do this we will require Stirling's formula, which says that

$$\Gamma(t) \ \sim \ e^{-t}t^t\sqrt{2\pi/t}.$$

So, we have that

$$\frac{\Gamma(k/2)}{\Gamma((k-1)/2)} \ \sim \ \frac{e^{-k/2}(k/2)^{k/2}}{e^{-(k-1)/2}((k-1)/2)^{(k-1)/2}} \ \sim \ \sqrt{k/2}.$$

Here we have used the fact that

$$\left(1-\frac{1}{k}\right)^k \ \sim \ 1/e, \text{ together with the fact that } \left(1-\frac{1}{k}\right)^c \ \sim \ 1,$$

for any fixed c (where $k \to \infty$).

So, for large k, the pdf for the Student's t distribution is

$$f(x) \ \sim \ \frac{1}{\sqrt{\pi}}\left(1+\frac{x^2}{k-1}\right)^{-k/2} \ \sim \ \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right).$$

Thus, as claimed, the Student's t distribution is approximately $N(0,1)$ as k tends to infinity.

3.2 Applying the Theorem to solve Problem 2

We seek δ so that $(\hat{\mu}-\delta_1, \hat{\mu}+\delta_1)$ contains μ at least 95% of the time. As we know, it turns out that this is the same as saying $\hat{\mu}$ lies in $(\mu-\delta, \mu+\delta)$ at least 95% of the time.

Now, we know that

$$t \ = \ \frac{(\hat{\mu}-\mu)\sqrt{k}}{\hat{\sigma}}$$

has a Student t distribution with $k-1$ degrees of freedom. Denote the cumulative distribution function for t by $\Psi(t)$.

To say that $\hat{\mu} \in (\mu-\delta, \mu+\delta)$ is the same as saying that

$$t \ \in \ \left(\frac{-\delta\sqrt{k}}{\hat{\sigma}}, \ \frac{\delta\sqrt{k}}{\hat{\sigma}}\right).$$

So, we seek δ so that

$$\Psi(\delta\sqrt{k}/\hat{\sigma}) - \Psi(-\delta\sqrt{k}/\hat{\sigma}) \ = \ 0.95.$$

As with $\Phi(t)$, the cdf for $N(0,1)$, we have that $\Psi(-t) = 1 - \Psi(t)$; and so, we seek δ so that

$$2\Psi(\delta\sqrt{k}/\hat{\sigma}) - 1 \ = \ 0.95.$$

That is,

$$\Psi(\delta\sqrt{k}/\hat{\sigma}) \ = \ 0.975.$$

This can easily be computed given tables of the Student t cumulative distribution values (recall, t is Student t with $k-1$ degrees of freedom).

3.3 A confidence interval for the variance

We will also determine a confidence interval for the variance, but first we need a bit of notation: We let $\chi^2_{\alpha,k}$ denote the αth *upper percentile* of a chi-squared random variable with k degrees of freedom, which means that if $f_k(x)$ is the pdf for χ^2_k, then

$$\int_{\chi^2_{\alpha,k}}^{\infty} f_k(x)dx \ = \ \alpha.$$

These values of $\chi^2_{\alpha,k}$ can be looked up in a table (or computed numerically using Maple, say).

Now, note that if $0 \leq a \leq b \leq 1$, then

$$\mathbb{P}(\chi^2_{b,k-1} \ \leq \ \chi^2_{k-1} \ \leq \ \chi^2_{a,k-1}) \ = \ b - a.$$

To see this, we observe that this probability is

$$\begin{aligned}
&\mathbb{P}(\chi^2_{k-1} \ \leq \ \chi^2_{a,k-1}) - \mathbb{P}(\chi^2_{k-1} \leq \chi^2_{b,k-1}) \\
&\quad = \ (1 - \mathbb{P}(\chi^2_{k-1} > \chi^2_{a,k-1})) - (1 - \mathbb{P}(\chi^2_{k-1} > \chi^2_{b,k-1}) \\
&\quad = \ \mathbb{P}(\chi^2_{k-1} \ > \ \chi^2_{b,k-1}) \ - \ \mathbb{P}(\chi^2_{k-1} \ > \ \chi^2_{a,k-1}) \\
&\quad = \ b - a.
\end{aligned}$$

So, when we go to use this to produce a probability p confidence interval, we will want that $b - a = p$. A good choice for a and b, to keep things nice and symmetric, is to simply take

$$a \ = \ (1-p)/2, \ b \ = \ (1+p)/2.$$

In the case $p = 0.95$ as we used earlier, this gives $a = 0.025$ and $b = 0.975$.

Now, as a consequence of the second part of Theorem 1, we have that

$$\mathbb{P}(\chi^2_{0.975,k-1} \ \leq \ \frac{(k-1)\hat{\sigma}^2}{\sigma^2} \ \leq \ \chi^2_{0.025,k-1}) \ = \ 0.95.$$

We want to turn this into a 95% confidence interval for σ^2, which will require rearranging things a little: We have that

$$\mathbb{P}(\frac{(k-1)\hat{\sigma}^2}{\sigma^2} \ \leq \ \chi^2_{0.025,k-1}) \ = \ \mathbb{P}(\sigma^2 \ \geq \ \frac{(k-1)\hat{\sigma}^2}{\chi^2_{0.025,k-1}});$$

and

$$\mathbb{P}(\chi^2_{0.975,k-1} \ \leq \ \frac{(k-1)\hat{\sigma}^2}{\sigma^2}) \ = \ \mathbb{P}(\sigma^2 \ \leq \ \frac{(k-1)\hat{\sigma}^2}{\chi^2_{0.975,k-1}}).$$

So, the event

$$\sigma^2 \ \in \ \left[\frac{(k-1)\hat{\sigma}^2}{\chi^2_{0.025,k-1}}, \ \frac{(k-1)\hat{\sigma}^2}{\chi^2_{0.975,k-1}}\right]$$

occurs with probability 0.95, and therefore this is a 95% confidence interval for σ^2.