Sampling Distributions, examples

CSU, Chico Math 314

2018-10-01

outline

Can we touch the CLT?

Assume $X_1, \ldots, X_n \sim_{iid} G(\theta)$. Simulate the sampling distribution of $T(x)$.

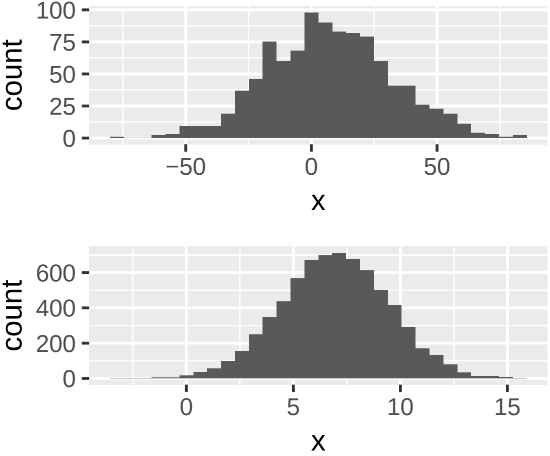
1. Randomly sample n observations from $G(\theta)$ – choose appropriate parameters.
2. Calculate and store $T_r(\mathbf{X})$.
3. Repeat steps 1 through 3 R times.
4. Make a histogram of $T_1, \ldots, T_R$.

Normal Example

```
R <- 7000
out <- rep(NA, R)
for(i in 1:R){
    ## assume population
    X <- rnorm(100, mean=7, sd=25)
    out[i] <- mean(X) # samples from sampling distribution
}
X <- rnorm(1001, 7, 25) # population
pop_plot <- ggplot(data.frame(x=X), aes(x=x)) +
    geom_histogram(bins=30)
sampdist_plot <- ggplot(data.frame(x=out), aes(x=x)) +
    geom_histogram(bins=30)
```

Normal Example

```
library(gridExtra)
grid.arrange(pop_plot, sampdist_plot, nrow=2)
```

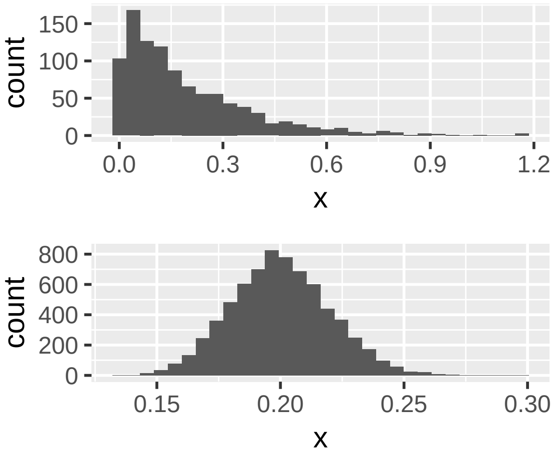


Exponential Example

```
R <- 7000
out <- rep(NA, R)
for(i in 1:R){
    ## assume population
    X <- rexp(100, 5)
    out[i] <- mean(X) # samples from sampling distribution
}
X <- rexp(1001, 5) # population
pop_plot <- ggplot(data.frame(x=X), aes(x=x)) +
    geom_histogram(bins=30)
sampdist_plot <- ggplot(data.frame(x=out), aes(x=x)) +
    geom_histogram(bins=30)
```

Exponential Example

```
grid.arrange(pop_plot, sampdist_plot, nrow=2)
```



Bias

Recall, bias is defined as $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$. How does one estimate $E(\hat{\theta})$?

Estimate Bias

Assume $X_1, \ldots, X_n \sim_{iid} Bernoulli(p)$. Estimate the bias of the estimator $\bar{X}$ – as calculated by `optim`. Steps:

1. Randomly sample n observations from $Bernoulli(p)$: `z <- rbinom(n, 1, p)`.
2. Estimate p.
3. Calculate and store bias: $bias_r = \hat{p} - p$
4. Repeat steps 1 through 3 R times and calculate $B = R^{-1} \sum_{r=1}^{R} bias_r$.

Bernoulli Example

```
pop_p <- 0.654
R <- 7000
out <- rep(NA, R)
for(i in 1:R){
    ## assume population
    X <- rbinom(100, 1, pop_p)
    out[i] <- mean(X) - pop_p # Point estimates of Bias,
}
mean(out) # the mean of which should be near zero.

## [1] -0.001315714
```

references I