C31 Parameter Estimation

David Housman

## Concepts

1. After obtaining data $x\_{1},x\_{2},…,x\_{n}$ from a sample of a population, we summarize with a statistic $\hat{θ}\left(x\_{1},x\_{2},…,x\_{n}\right)$ such as the mean, median, standard deviation, or interquartile range. This is part of descriptive statistics.
2. We hope that the statistic $\hat{θ}$ tells us something about a population parameter $θ$ such as $λ$ in the Poisson random variable, $α$ and $β$ in the gamma random variable, or $p$ in the geometric random variable. In this context, the statistic $\hat{θ}$ is called an *estimator* for the population parameter $θ$. This is part of inferential statistics.
3. Right now, we focus on probability. We start with a known population, modeled by a random variable $X$ governed by a parameter $θ$; take a sample, modeled by mutually independent random variables $X\_{1},X\_{2},…,X\_{n}$ each having the same distribution as $X$; and study the *statistic*
* $  \hat{θ}\left(X\_{1},X\_{2},…,X\_{n}\right)$
* and its distribution (usually called a *sampling distribution*). Observe our use of capital letters when referring to random variables and lower case letters with referring to data.
1. We hope that our statistics are unbiased (a statistic’s expected value $E\left[\hat{θ}\left(X\_{1},X\_{2},…,X\_{n}\right)\right]$ equals the parameter $θ$ being estimated) and are precise (a statistic’s variance $Var\left[\hat{θ}\left(X\_{1},X\_{2},…,X\_{n}\right)\right]$ is as small as possible).
2. Recall that if $x$ and $Y$ are random variables and $a$ and $b$ are real numbers, then $E\left[aX+bY\right]=aE\left[X\right]+bE\left[Y\right]$ and $Var\left[aX+bY\right]=a^{2}Var\left[X\right]+b^{2}Var\left[Y\right]+2abCov\left(X,Y\right)$. If $X$ and $Y$ are also independent, then $Var\left[aX+bY\right]=a^{2}Var\left[X\right]+b^{2}Var\left[Y\right]$.
3. Suppose x = c(1,2,3). Then rep(x, times = 4) returns c(1,2,3,1,2,3,1,2,3,1,2,3), and rep(x, each = 4) returns c(1,1,1,1,2,2,2,2,3,3,3,3).

## Poisson Population Example

In an experiment conducted by Rutherford, Chadwick, and Ellis, a radioactive substance was observed during 2,608 time intervals of 7.5 seconds each; the number of particles reaching a counter was obtained for each period. The sample mean and variance were 3.87 and 3.68 particles, respectively, and we wondered whether this data could have arisen from a Poisson distribution with mean around 3.87. This is an inferential statistics question. Let us answer a related probability question: For a population modeled by a Poisson distribution with $λ=3.87$, what would samples of size 2,608 look like?

1. Generate a sample of size 2,608 from a Poisson random variable with parameter $λ=3.87$. Describe the sample with a histogram, mean, and variance. Repeat a few times to get some intuition about the amount of variation in the sample histograms, means, and variances.

n = 2608
lambda = 3.87
sample = tibble(x = rpois(n, lambda))
ggplot(sample, aes(x = x)) + geom\_histogram(binwidth = 1, center = 0, color = "white")



sample |> summarise(mean(x), var(x))

# A tibble: 1 × 2
 `mean(x)` `var(x)`
 <dbl> <dbl>
1 3.87 3.82

1. Simulate 10,000 samples of size 2,608 from a Poisson distribution with $λ=3.87$. Save in a tibble named samples with columns run and x.

nruns = 10000
n = 2608
lambda = 3.87
samples = tibble(
 run = rep(1:nruns, each = n),
 x = rpois(nruns\*n, lambda))

## Sample Mean Statistic $‾$

Suppose $X\_{1},X\_{2},…,X\_{n}$ are independent identically distributed random variables each having mean $μ$ and standard deviation $σ$. Let the sample mean statistic

$$‾=\frac{1}{n}\left(X\_{1}+X\_{2}+\cdots +X\_{n}\right)$$

be an estimator for the population mean $μ$.

1. Calculate the mean for each sample in samples. Display a histogram of the sample means. Calculate the mean of the sample means, the standard deviation of the sample means, and the interval within which 95% of the sample means lie. Write a sentence summarizing the results. How might this provide an inference from the original data?

means = samples |> group\_by(run) |> summarise(xbar = mean(x))

ggplot(means, aes(x = xbar)) +
 geom\_histogram()

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



means |> summarise(
 mean = mean(xbar),
 sd = sd(xbar),
 lo = quantile(xbar, 0.025),
 hi = quantile(xbar, 0.975)
)

# A tibble: 1 × 4
 mean sd lo hi
 <dbl> <dbl> <dbl> <dbl>
1 3.87 0.0390 3.79 3.95

The distribution of sample means looks bell shaped with a mean of 3.87, standard deviation of 0.04, and a 95% coverage interval of 3.79 to 3.95.

1. Determine the mean, variance, standard deviation, and distribution of the random variable $‾$ in general. Apply these results when each $X\_{i}$ is $Poisson\left(3.87\right)$. How is this related to the previous exercise?

The mean of $‾$ is

$$\begin{matrix}μ\_{‾}&=&E\left[‾\right]\\&=&E\left[\frac{1}{n}\left(X\_{1}+X\_{2}+\cdots +X\_{n}\right)\right]\\&=&\frac{1}{n}\left(E\left[X\_{1}\right]+E\left[X\_{2}\right]+\cdots +E\left[X\_{n}\right]\right)\\&=&\frac{1}{n}\left(μ+μ+\cdots +μ\right)\\&=&μ.\end{matrix}$$

The variance of $‾$ (using independence) is

$$\begin{matrix}Var\left[‾\right]&=&Var\left[\frac{1}{n}\left(X\_{1}+X\_{2}+\cdots +X\_{n}\right)\right]\\&=&\frac{1}{n^{2}}\left(Var\left[X\_{1}\right]+Var\left[X\_{2}\right]+\cdots +Var\left[X\_{n}\right]\right)\\&=&\frac{1}{n^{2}}\left(σ^{2}+σ^{2}+\cdots +σ^{2}\right)\\&=&\frac{1}{n}σ^{2}.\end{matrix}$$

The standard deviation is

$  σ\_{‾}=\sqrt{Var\left[‾\right]}=\sqrt{\frac{1}{n}σ^{2}}=\frac{σ}{\sqrt{n}}.$

The Central Limit Theorem tells us that the distribution of $‾$ is approximately normal for sufficiently large $n$.

For $X\_{i}∼Poisson\left(3.87\right)$ and $n=2608$, $μ\_{‾}=3.87$ and $σ\_{‾}=\sqrt{3.87/2608}=0.0385$, and the central 95% of the sample means should be in 3.87 ± 0.08 particles per period. This matches what we found in the previous exercise because that was a simulation of $‾$.

means |> summarise(
 smean = mean(xbar),
 tmean = lambda,
 ssd = sd(xbar),
 tsd = sqrt(lambda/n)
)

# A tibble: 1 × 4
 smean tmean ssd tsd
 <dbl> <dbl> <dbl> <dbl>
1 3.87 3.87 0.0390 0.0385

ggplot(means, aes(x = xbar)) +
 geom\_histogram(aes(y = after\_stat(density))) +
 geom\_function(fun = dnorm, args = list(mean = lambda, sd = sqrt(lambda/n)), color = "red", linewidth = 2)

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



1. What do the results from the previous exercise tell us about the use of the statistic $‾$ to estimate the parameter $μ$ of a population?

## Sample Variance Statistic $S^{2}$

Suppose $X\_{1},X\_{2},…,X\_{n}$ are independent identically distributed random variables each having mean $μ$ and standard deviation $σ$. Let the sample variance statistic

$$S^{2}=\frac{1}{n−1}\left(\left(X\_{1}−‾\right)^{2}+\left(X\_{2}−‾\right)^{2}+\cdots \left(X\_{n}−‾\right)^{2}\right)$$

be an estimate for the population variance $σ^{2}$.

1. Calculate the variance for each sample in samples. Display a histogram of the sample variances. Calculate the mean of the sample variances, the standard deviation of the sample variances, and the interval within which 95% of the sample variances lie. Write a sentence summarizing the results. How might this provide an inference from the original data?

vars = samples |> group\_by(run) |> summarise(s2 = var(x))

ggplot(vars, aes(x = s2)) +
 geom\_histogram()

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



vars |> summarise(
 mean = mean(s2),
 sd = sd(s2),
 lo = quantile(s2, 0.025),
 hi = quantile(s2, 0.975)
)

# A tibble: 1 × 4
 mean sd lo hi
 <dbl> <dbl> <dbl> <dbl>
1 3.87 0.113 3.65 4.09

1. Determine the mean, variance, standard deviation, and distribution of the random variable $S^{2}$. Apply these results when each $X\_{i}$ is $Poisson\left(3.87\right)$. How is this related to the previous exercise?

Consider

$$\begin{matrix}E\left[\left(n−1\right)S^{2}\right]&=&E\left[\sum\_{i=1}^{n}\left(X\_{i}−‾\right)^{2}\right]\\&=&E\left[\sum\_{i=1}^{n}\left(\left(X\_{i}−μ\right)−\left(‾−μ\right)\right)^{2}\right]\\&=&E\left[\sum\_{i=1}^{n}\left(\left(X\_{i}−μ\right)^{2}−2\left(X\_{i}−μ\right)\left(‾−μ\right)+\left(‾−μ\right)^{2}\right)\right]\\&=&E\left[\sum\_{i=1}^{n}\left(\left(X\_{i}−μ\right)^{2}−2\left(X\_{i}−μ\right)\frac{1}{n}\sum\_{j=1}^{n}\left(X\_{j}−μ\right)+\left(‾−μ\right)^{2}\right)\right]\\&=&\sum\_{i=1}^{n}E\left[\left(X\_{i}−μ\right)^{2}\right]−\frac{2}{n}\sum\_{i=1}^{n}\sum\_{j=1}^{n}E\left[\left(X\_{i}−μ\right)\left(X\_{j}−μ\right)\right]+\sum\_{i=1}^{n}E\left[\left(‾−μ\right)^{2}\right]\\&=&nσ^{2}−\frac{2}{n}nσ^{2}+n\frac{σ^{2}}{n}\\&=&\left(n−1\right)σ^{2}\end{matrix}$$

Thus, the mean of $S^{2}$ is

$  E\left[S^{2}\right]=σ^{2}$.

It can be shown (https://math.stackexchange.com/questions/72975/variance-of-sample-variance) that

$  Var\left(S^{2}\right)=\frac{μ\_{4}}{n}−\frac{σ^{4}\left(n−3\right)}{n\left(n−1\right)}$

where $μ\_{4}=E\left[\left(X−μ\right)^{4}\right]$ is the fourth central moment of $X$.

A corollary to the Central Limit Theorem is that the distribution of $S^{2}$ is approximately normal for large sample sizes (https://math.stackexchange.com/questions/2325500/central-limit-theorem-for-the-variance#mjx-eqn-2).

For $X\_{i}∼Poisson\left(3.87\right)$ (the s values arise from our simulation and the t values arise from the above theory):

vars |> summarise(
 smean = mean(s2),
 tmean = lambda,
 ssd = sd(s2),
 tsd = sqrt((3\*lambda^2+lambda)/n - lambda^2\*(n-3)/n/(n-1))
)

# A tibble: 1 × 4
 smean tmean ssd tsd
 <dbl> <dbl> <dbl> <dbl>
1 3.87 3.87 0.113 0.114

The histogram is of the simulated $S^{2}$ and the red curve is of the theoretical pdf for $S^{2}$:

ggplot(vars, aes(x = s2)) +
 geom\_histogram(aes(y = after\_stat(density))) +
 geom\_function(fun = dnorm, args = list(mean = lambda, sd = sqrt((3\*lambda^2+lambda)/n - lambda^2\*(n-3)/n/(n-1))), color = "red", linewidth = 2)

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



1. What do the results from the previous exercise tell us about the use of the statistic $S^{2}$ to estimate the parameter $σ^{2}$ of a population?
2. What do the results from exercises 4 and 7 tell us about the use of the statistics $‾$ and $S^{2}$ to estimate the parameter $λ$ of a Poisson population?

## Finding Estimators

Two approaches to finding a statistic to estimate a parameter are the method of moments and maximizing likelihood. We illustrate these approaches via some examples.

1. Find the method of moments estimators for the mean $μ$ and variance $σ^{2}$ of a normal population.

The central moments of a random variable $X$ are $μ\_{1}=E\left[X\right]$ and $μ\_{k}=E\left[\left(X−μ\_{1}\right)^{k}\right]$ for positive integers $k$.

If $X$ is randomly selecting an element from a list of data $x\_{1},x\_{2},…,x\_{n}$, then $μ\_{1}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}=‾$ and $μ\_{2}=\frac{1}{n}\sum\_{i=1}^{n}\left(x\_{i}−‾\right)^{2}=\frac{n−1}{n}s^{2}=v$.

If $X$ is a normal distribution with mean $μ$ and standard deviation $σ$, then $μ\_{1}=μ$ and $μ\_{2}=σ^{2}$.

Matching the first two central moments yields $\hat{μ}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}$ and $\hat{σ^{2}}=\frac{1}{n}\sum\_{i=1}^{n}\left(x\_{i}−‾\right)^{2}$.

Note that $\hat{μ}$ is the sample mean but $\hat{σ^{2}}$ is the natural but biased $\frac{n−1}{n}$ times the sample variance. Note also that neither $S$ nor $\sqrt{V}$ is an unbiased estimator for $σ$.

1. Find the method of moments estimator for the mean $λ$ of a Poisson population.

If $X$ is a Poisson distribution with mean $λ$, then $μ\_{1}=λ$. Matching the first central moments yields $\hat{λ}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}$.

1. Find the maximum likelihood estimator for the mean $λ$ of a Poisson population.

Our population $X$ has a pdf of the form $f\left(x\right)=e^{−λ}\frac{λ^{x}}{x!}$ for $x=0,1,2,…$ where $λ>0$ is unknown. We obtain data $x\_{1},x\_{2},…,x\_{n}$ from a sample of the population and endeavor to estimate the parameter $λ$. The probability (called likelihood in this context) that a sample from the given population would be the same as our actual data is

$  L\left(λ\right)=\prod\_{i=1}^{n}f\left(x\_{i}\right)=\prod\_{i=1}^{n}e^{−λ}\frac{λ^{x\_{i}}}{x\_{i}!}$.

We want to choose our parameter $λ$ so as to maximize the likelihood $L\left(λ\right)$. It is often easier to maximize the log likelihood:

$$\begin{matrix}l\left(λ\right)&=&ln\left(L\left(λ\right)\right)\\&=&\sum\_{i=1}^{n}\left(ln\left(e^{−λ}\right)+ln\left(λ^{x\_{i}}\right)−ln\left(x\_{i}!\right)\right)\\&=&−nλ+ln\left(λ\right)\sum\_{i=1}^{n}x\_{i}−\sum\_{i=1}^{n}ln\left(x\_{i}!\right)\end{matrix}$$

A necessary condition to maximize $l$ comes from calculus:

$  0=l′\left(\hat{λ}\right)=−n+\frac{1}{\hat{λ}}\sum\_{i=1}^{n}x\_{i}$

The unique solution is

$  \hat{λ}=\frac{1}{n}\sum\_{i=1}^{n}x\_{i}$.

It is easy to verify with a first derivative test that this is a global maximum.

## Gamma Population

Suppose we have a gamma population with unknown parameters $α$ and $β$. We obtain a sample $x\_{1},x\_{2},…,x\_{n}$ and want to estimate the population parameters.

1. Find the method of moments estimators.

Recall from *C25 Important Random Variables* that for $Gamma\left(α,β\right)$, the mean is $μ=αβ$ and the variance is $σ^{2}=αβ^{2}$. By solving these two equations for the parameters, we obtain $β=σ^{2}/μ$ and $α=μ^{2}/σ^{2}$. Thus, the method of moments estimators are $\hat{α}=‾^{2}/v$ and $\hat{β}=v/‾$ where $v=\frac{n−1}{n}s^{2}$.

1. Find the maximum likelihood estimators.

Recall from *C25 Important Random Variables* that for $Gamma\left(α,β\right)$, the pdf is $f\left(x\right)=\frac{1}{Γ\left(α\right)β^{α}}x^{α−1}e^{−x/β}$ for $x\geq 0$ where $Γ\left(α\right)=\int\_{0}^{\infty }t^{α−1}e^{−t}dt$ is a generalization of $\left(α−1\right)!$. Thus, the log likelihood function is

$$\begin{matrix}l\left(α,β\right)&=&\sum\_{i=1}^{n}ln\left(\frac{1}{Γ\left(α\right)β^{α}}x\_{i}^{α−1}e^{−x\_{i}/β}\right)\\&=&−nln\left(Γ\left(α\right)\right)−nαln\left(β\right)+\left(α−1\right)\left(\sum\_{i=1}^{n}ln\left(x\_{i}\right)\right)−\frac{1}{β}\sum\_{i=1}^{n}x\_{i}\end{matrix}$$

Necessary conditions to maximize $l$ are

$  0=\frac{∂l}{∂α}=−n\frac{Γ′\left(α\right)}{Γ\left(α\right)}−nln\left(β\right)+\sum\_{i=1}^{n}ln\left(x\_{i}\right)$

and

$  0=\frac{∂l}{∂β}=−n\frac{α}{β}+\frac{1}{β^{2}}\sum\_{i=1}^{n}x\_{i}$.

The second equation yields

$  \hat{α}\hat{β}=‾$

but the first equation is not very helpful symbolically. We must find a numerical solution for any given list of data.

1. Suppose the data given below came from a gamma population. Find the method of moments estimators and maximum likelihood estimators. Compare the empirical cdf and the two theoretical cdfs in a single plot.

actual = tibble(x = c(49.2,53.9,50.0,44.5,42.2,42.3,32.3,31.3,60.9,47.5))

moments = actual |> summarize(
 n = n(),
 xbar = mean(x),
 v = (n-1)/n \* var(x),
 alpha = xbar^2 / v,
 beta = v / xbar
)
moments

# A tibble: 1 × 5
 n xbar v alpha beta
 <int> <dbl> <dbl> <dbl> <dbl>
1 10 45.4 74.4 27.7 1.64

likelihood = actual |> summarize(
 n = n(),
 xbar = mean(x),
 v = (n-1)/n \* var(x),
 alpha = uniroot(function(a) -n\*digamma(a) - n\*log(xbar/a) + sum(log(x)), c(10, 40))$root,
 beta = xbar / alpha
)
likelihood

# A tibble: 1 × 5
 n xbar v alpha beta
 <int> <dbl> <dbl> <dbl> <dbl>
1 10 45.4 74.4 26.4 1.72

ggplot(actual, aes(x = x)) +
 stat\_ecdf() +
 geom\_function(fun = pgamma, args = list(shape = moments$alpha, scale = moments$beta), xlim = c(29, 63), color = "red") +
 geom\_function(fun = pgamma, args = list(shape = likelihood$alpha, scale = likelihood$beta), xlim = c(29, 63), color = "blue")



It appears that the data could have come from a gamma population and that the two theoretical distributions are quite similar.