

Statistics CH1

Xiang Li

2023/10/18

1.2.0.1 Exercise

Suppose X has a Poisson distribution with λ . If $\lambda=4$, what is the probability that we observe the values $X=1,3,4,2$ from this distribution?

```
x = c(1,3,4,2)
p = dpois(x, lambda = 4)
prod(p)
```

```
## [1] 0.0004097281
```

1.2.0.2 Exercise

Suppose X has a Poisson distribution with λ . If we observe $X=1,3,4,2$. Make a guess of the value of λ do you think could have been the one that generated the data X ? Motivate your guess.

```
lamb_v = c(1.5, 2, 2.5, 3, 3.5)
prod_p = function(lamb) {
  x = c(1,3,4,2)
  return(prod(dpois(x, lambda = lamb)))
}
prod_p_v = sapply(lamb_v, prod_p)
names(prod_p_v) = lamb_v
prod_p_v
```

```
##          1.5          2          2.5          3          3.5
## 0.0004963102 0.0011927560 0.0015033593 0.0012597555 0.0007964623
```

1.2.0.3 Exercise

Explain in your own words why statisticians are more interested in the underlying probability distribution, rather than in the data themselves?

The underlying probability distribution can describe more general cases rather than special cases which only fit this data set.

1.3.0.1 Exercise

In which of the following situations, what are the sampling units. Is it reasonable to assume that the sample is iid? What arguments can you give against it?

1. The saliva of all attendants at a dance club is tested for drugs.
2. I measure 24h precipitation every Monday at the same location for a year
3. *Sample unit is one person's saliva. iid.*

Not iid. The attendants may come in groups, and they may have influenced each others drug use. The probability that one will use drug is not same distributed. 2. Sample unit is one Monday's precipitation. Not iid. The precipitation in different seasons have different distributions and the precipitation at adjacent times are correlated.

1.3.0.2 Exercise

The binomial and the Poisson distribution are both used when the data are a count. Explain when to use a binomial and when to use a Poisson distribution.

The binomial distribution describes the sum of n variables that follow Bernoulli distribution, i.e. the results of n trials. The Poisson distribution describes the number of times an event occurs in a given interval of time.

Binomial distributions have a known parameter n . They are used when the maximal count (n) is fixed and known. Poisson distributions are used when the counts have no (known) maximum.

1.3.0.3 Exercise

What model do you think could be appropriate for data where the following was measured for each sampling unit. Motivate!

1. The time between graduating and finding a job
2. Concentration of a hormone in blood
3. The difference between the scores of two consecutive IQ tests
4. The number of times in a year a student's bicycle has a flat tire
5. *Exponential distribution.*
6. *2. Normal distribution.*
7. *3. Normal distribution.*
8. *4. Poisson distribution.*

1.3.0.4 Exercise

Explain the quote "All models are wrong, but some are useful." in your own words. Do you agree?

Agree. We can't find the exact model which could describe the population totally, but we can find the model which is most similar with the exact one.

1.4.0.1 Exercise

Recall that if $X \sim \text{Poisson}(\lambda)$, then $\bar{X} \approx E(X) = \lambda$.

Now, suppose X has a Poisson distribution with parameter λ . If we observe $X=1,3,4,2$, what is a method of moments estimate of λ ?

```
x = c(1,3,4,2)
lambda_hat = mean(x)
lambda_hat
```

```
## [1] 2.5
```

1.4.0.2 Exercise

Suppose X has an exponential distribution with rate parameter λ . We observe $X=(0.3,0.6,6.4,0.5,0.8,0.2)$. What is a method-of-moments estimate for λ ?

```
x = c(0.3,0.6,6.4,0.5,0.8,0.2)
lambda_hat = 1/mean(x)
lambda_hat
```

```
## [1] 0.6818182
```

1.4.0.3 Exercise

Suppose X has a binomial distribution with parameters $n=8$ and p . We observe $X=(7,7,5,6,6,3,7,7,4,6)$. What is a method-of-moments estimate for p ?

```
x = c(7,7,5,6,6,3,7,7,4,6)
n = 8
p_hat = mean(x)/n
p_hat
```

```
## [1] 0.725
```

1.4.0.4 Exercise

Suppose X has a uniform distribution with parameters a and θ . We observe $X=(0.1,0.2,0.4,1.5)$. What is a method-of-moments estimate for θ ? Is this a good estimate? Why?

```
x = c(0.1,0.2,0.4,1.5)
a = 0
theta_hat = 2*mean(x) - a
theta_hat
```

```
## [1] 1.1
```

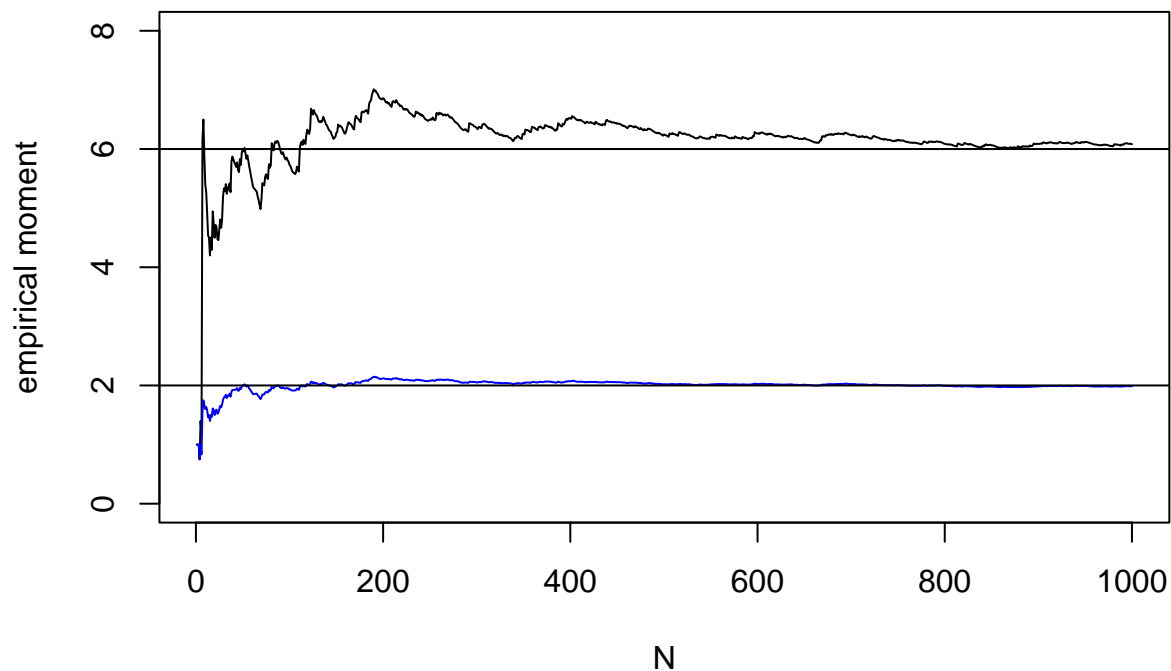
This is not a good estimate. Because we observe 1.5, θ should be greater than or equal to 1.5.

1.4.0.5 Exercise

We know from the law of large numbers that $\frac{1}{n} \sum_{i=1}^n X_i$ converges to $E(X)$ as n grows. Why does it also follow from the law of large numbers that $\frac{1}{n} \sum_{i=1}^n X_i^k$ converges to $E(X^k)$? Explain.

See X_i^k as a random variable and then use law of large number.

```
X = rpois(1000, lambda=2)
N = 1:1000
mom1 = mom2 = numeric(1000)
for (i in 1:1000) {
  mom1[i] <- mean(X[1:i])
  mom2[i] <- mean(X[1:i]^2)
}
plot(N, mom2, type='l', ylim=c(0,8), ylab='empirical moment')
lines(N, mom1, col='blue')
abline(h=c(2,6))
```



1.4.0.6 Exercise Which of the two empirical moments is typically closer to its expectation? Can you give an intuitive explanation for this?

The 1st moment estimator is closer to its expectation. X^2 gives extremely large values more often than X itself. Such outliers make the estimate variable and take more time to average out.

1.4.0.7 Exercise

Show that $Var(X) = \mu_2 - \mu^2$.

$$Var(X) = E[X - E(X)]^2 = E[X^2 - 2XE(X) + E^2(X)] = E(X^2) - E^2(X) = \mu_2 - \mu^2$$

1.4.0.8 Exercise

Show that $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \hat{\mu}_2 - \hat{\mu}^2$.

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}\bar{X} + \bar{X}^2 = \hat{\mu}_2 - \hat{\mu}^2$$

1.4.0.9 Exercise

Suppose X has a Poisson distribution with parameter λ . For X=1,3,4,2, find a method of moments estimate of λ using the property that the variance of the Poisson distribution is λ . Do you think that this is a better estimate than the one based on the mean?

```
X = c(1,3,4,2)
lambda_hat1 = mean(X)
lambda_hat2 = var(X)*(length(X)-1)/length(X)
lamb_v = c(lambda_hat1, lambda_hat2)
prod_p = function (lamb) {
  x = c(1,3,4,2)
  return(prod(dpois(x, lambda = lamb)))
}
prod_p_v = sapply(lamb_v, prod_p)
names(prod_p_v) = lamb_v
prod_p_v
```

```
##          2.5          1.25
## 0.001503359 0.000217889
```

The estimate based on the mean is better.

1.4.0.10 Exercise

Suppose X=(1.7,2.0,3.2,0.7,5.4,1.0) are observed from a gamma-distribution with parameters α and β . We can look up the gamma distribution and find that $E(X) = \alpha/\beta$ and $Var(X) = \alpha/\beta^2$. Find method-of-moments estimators for α and β .

```
X = c(1.7,2.0,3.2,0.7,5.4,1.0)
n = length(X)
alpha_hat = mean(X)**2/(var(X)*(n-1)/n)
beta_hat = mean(X)/(var(X)*(n-1)/n)
alpha_hat
```

```
## [1] 2.161447
```

```
beta_hat
```

```
## [1] 0.9263344
```

1.4.0.11 Exercise

We have observed $X=(0.40,0.23,1.83,2.24,1.61)$, which we assume to have come from a lognormal distribution with parameters v and τ^2 , with $\tau^2 = 1$ known. For the lognormal distribution, the expectation is $e^{v+\tau^2/2}$. Find a method-of-moments estimator based on this expression.

```
X = c(0.40,0.23,1.83,2.24,1.61)
tau_squ = 1
upsilon_hat = log(mean(X)) - tau_squ/2
upsilon_hat
```

```
## [1] -0.2673022
```

1.4.0.12 Exercise

We have again observed $X=(0.40,0.23,1.83,2.24,1.61)$, which we assume to have come from a lognormal distribution with parameters v and τ^2 , with $\tau^2 = 1$ known. The variance of a lognormal distribution is $(e^{\tau^2} - 1)e^{2v+\tau^2}$. Derive a method-of-moments estimate for v based on this expression.

```
X = c(0.40,0.23,1.83,2.24,1.61)
tau_squ = 1
n = length(X)
upsilon_hat1 = (log(var(X)*(n-1)/n/(exp(tau_squ)-1))-tau_squ)/2
upsilon_hat1
```

```
## [1] -0.9925139
```

1.4.0.14 Exercise

Suppose, again, that X is log-normally distributed with parameters v and τ^2 . We have observed $X=(0.40,0.23,1.83,2.24,1.61)$. It is a property of the lognormal distribution that if X is log-normally distributed with parameters v and τ^2 , then $Y = \log(X)$ is normally distributed with parameters $\mu = v$ and $\sigma^2 = \tau^2$. 1. Transform X to Y and formulate a parametric model for Y . 2. Find a method-of-moments estimator for $\mu = v$.

```
X = c(0.40,0.23,1.83,2.24,1.61)
Y = log(X)
Y
```

```
## [1] -0.9162907 -1.4696760 0.6043160 0.8064759 0.4762342
```

```
mu_hat = mean(Y)
mu_hat
```

```
## [1] -0.09978814
```

1.5.0.2 Exercise

Simulate data a couple of times to see the estimate vary. Now change the sample size to 100. What changes?

```
p_v1 = replicate(10, mean(rbinom(10, size=1, prob=0.42)))
p_v1
```

```
## [1] 0.3 0.6 0.3 0.5 0.4 0.1 0.5 0.4 0.4 0.0
```

```
p_v2 = replicate(10, mean(rbinom(100, size=1, prob=0.42)))
p_v2
```

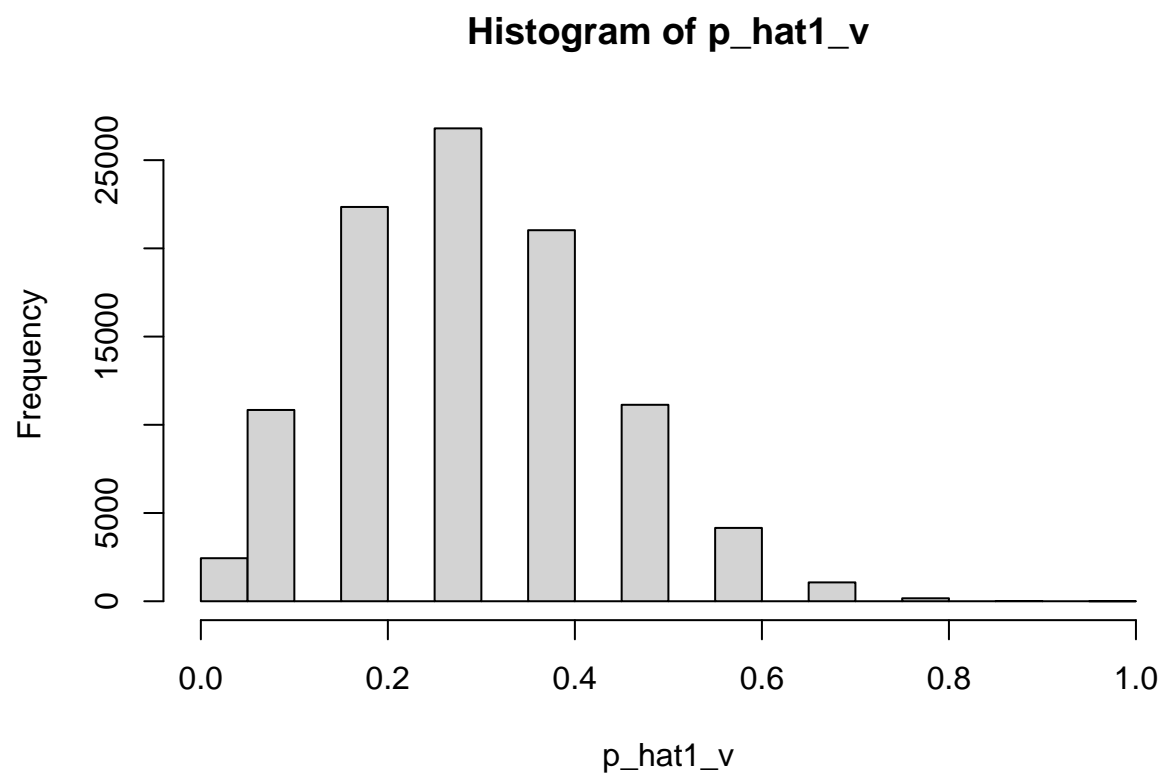
```
## [1] 0.41 0.38 0.42 0.38 0.42 0.37 0.44 0.34 0.52 0.44
```

The estimate of 100-sized sample is closer to the true value.

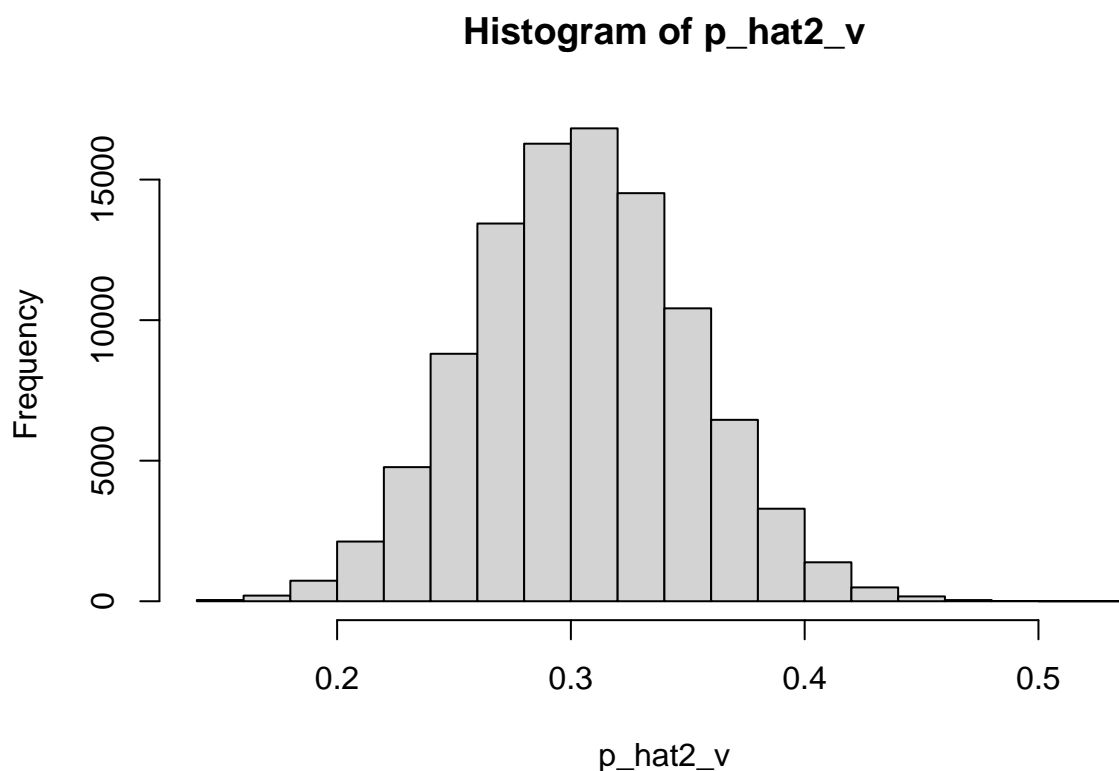
1.5.0.3 Exercise

Calculate the sampling distribution for the same estimator if the sample size is 10 and if the sample size is 100, and the true probability of success is 0.31. What is the probability that the estimator is larger than 0.40 in both cases?

```
p_hat1_v = replicate(1e5, {
  X = rbinom(10, size = 1, prob = 0.31)
  mean(X)
})
p_hat2_v = replicate(1e5, {
  X = rbinom(100, size = 1, prob = 0.31)
  mean(X)
})
hist(p_hat1_v)
```



```
hist(p_hat2_v)
```

```
result_p_hat1 = mean(p_hat1_v > 0.4)
result_p_hat2 = mean(p_hat2_v > 0.4)
result_p_hat1
```

```
## [1] 0.1654
```

```
result_p_hat2
```

```
## [1] 0.02105
```

While we often think of estimating as finding one or more parameters, we are essentially always estimating the probability distribution of the data.

```
mu = 5      # true mu
sigma = 2   # true sigma
x <- seq(0, 10, 0.01)
# true distribution
plot(x, dnorm(x, mu, sigma), type='l', ylim=c(0,.4), lwd=3)

estimates = replicate(20,
{
  X = rnorm(10, mu, sigma) # data
  hat.mu = mean(X)         # estimated mu
  hat.sig = sd(X)          # estimated sigma
})
```

```
# estimated distributions  
lines(x, dnorm(x, hat.mu, hat.sig), col="red", lwd=0.5)  
})
```

