Chapter 8

Grinstead

Jai Jeffryes

5/2/2020

Ex. 8.1.1

A fair coin is tossed 100 times. The expected number of heads is 50, and the standard deviation for the number of heads is $(100 \cdot 1/2 \cdot 1/2)1/2 = 5$. What does Chebyshev's Inequality tell you about the probability that the number of heads that turn up deviates from the expected number 50 by three or more standard deviations (i.e., by at least 15)?

Note:

- What's given about standard deviation? We know the number of trials is 100 and the probability of the outcome heads. The formula for calculating standard deviation for the binomial distribution is $\sigma = \sqrt{npq}$.
- The question asks for *at least* 3 standard deviations. Use the tail form of the theorem.

$P(|X - \mu| \geq k\sigma) \geq \frac{1}{k^2}$

So, $P(|X - 50| \geq 3 \cdot 5) >= \frac{1}{9} = 0.1111111$.

The probability that the number of heads is at least 15 is, at most, 0.1111111.

Ex. 8.1.2

Write a program that uses the function binomial(n, p, x) to compute the exact probability that you estimated in Exercise 1. Compare the two results.

I wrote this function to visualize and calculate all the binomial intervals when I reviewed the binomial distribution section for DATA606. *Note:* there still might be some work to do on some of the boundaries.

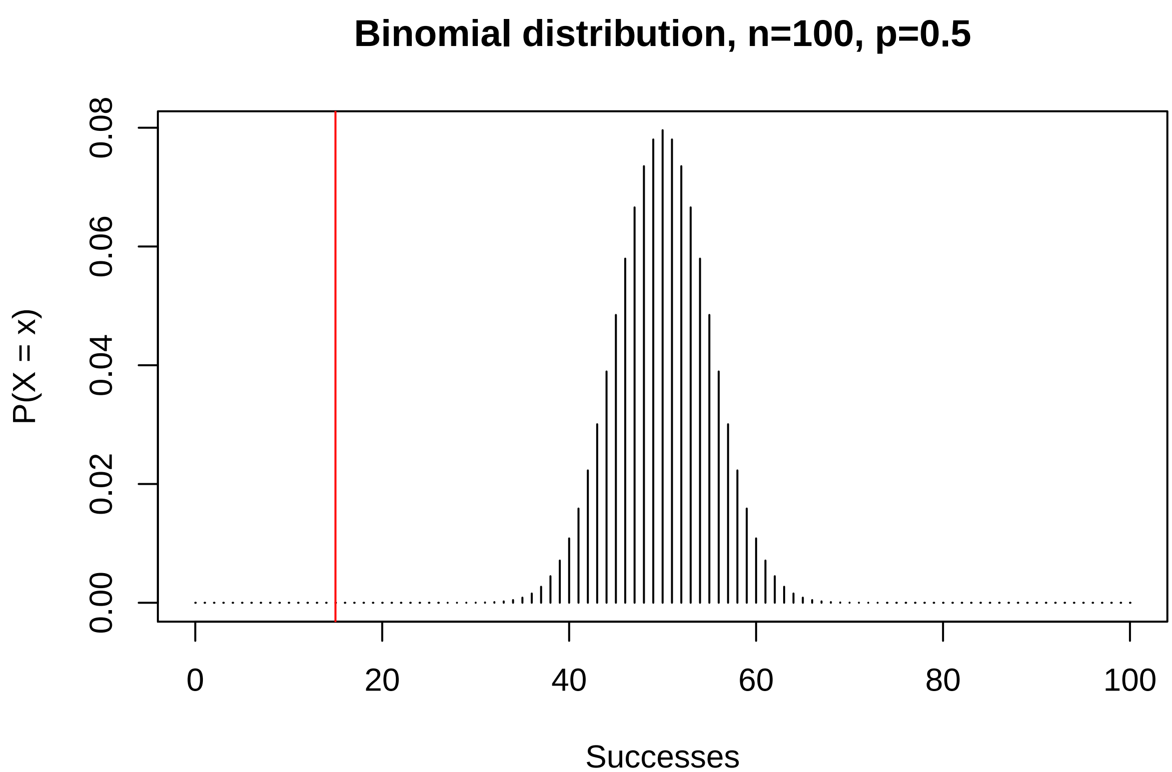
```
binom_summary <- function(successes, trials, prob,
                          main = NULL, xlab = "Successes", ylab = "P(X = x)") {
  # Annotate plot.
  param_str <- paste0(", n=", trials, ", p=", prob)
  main <- ifelse(is.null(main), paste0("Binomial distribution", param_str),
                 paste0(main, param_str))

  plot(0:trials, dbinom(0:trials, trials, prob = prob),
     xlab = xlab,
     ylab = ylab,
     type = "h",
     main = main)
  abline(v = successes, col = "red")

  mu <- trials * prob
  sigma <- sqrt(trials * prob * (1 - prob))
```

```r
is_normal_approx <- trials * prob >= 10 & trials * (1 - prob) >= 10

print(paste0("mu: ", mu))
print(paste0("sigma: ", sigma))

print(paste0("Trials: ", trials))
print(paste0("Successes: ", successes))

prob.exact <- dbinom(successes, trials, prob)

prob.cumulative.lt <- pbinom(successes - 1, trials, prob)
# Normal approximation
prob.cumulative.lt.approx <- pnorm(successes, mu, sigma)
# Normal approximation with continuity correction.
prob.cumulative.lt.approx.corr <- pnorm(successes - 0.5, mu, sigma)

prob.cumulative.le <- pbinom(successes, trials, prob)
# Normal approximation
prob.cumulative.le.approx <- pnorm(successes, mu, sigma)
# Normal approximation with continuity correction.
prob.cumulative.le.approx.corr <- pnorm(successes + 0.5, mu, sigma)

prob.cumulative.gt <- pbinom(successes, trials, prob, lower.tail = F)
# Normal approximation
prob.cumulative.gt.approx <- pnorm(successes, mu, sigma, lower.tail = F)
# Normal approximation with continuity correction.
prob.cumulative.gt.approx.corr <- pnorm(successes + 0.5, mu, sigma, lower.tail = F)

prob.cumulative.ge <- pbinom(successes - 1, trials, prob, lower.tail = F)
# Normal approximation
prob.cumulative.ge.approx <- pnorm(successes, mu, sigma, lower.tail = F)
# Normal approximation with continuity correction.
prob.cumulative.ge.approx.corr <- pnorm(successes - 0.5, mu, sigma, lower.tail = F)

print(paste0("Binomial probability. P(X=x): ", round(prob.exact, 4)))

print(paste0("Cumulative probability. P(X<x): ", round(prob.cumulative.lt, 4)))
if (is_normal_approx) {
  print(paste0("    Normal approximation. P(X<x): ",
               round(prob.cumulative.lt.approx, 4)))
  print(paste0("    Normal approximation w/ continuity correction. P(<x): ",
               round(prob.cumulative.lt.approx.corr, 4)))
}

print(paste0("Cumulative probability. P(X<=x): ", round(prob.cumulative.le, 4)))
if (is_normal_approx) {
  print(paste0("    Normal approximation. P(X<=x): ",
               round(prob.cumulative.le.approx, 4)))
  print(paste0("    Normal approximation w/ continuity correction. P(<=x): ",
               round(prob.cumulative.le.approx.corr, 4)))
}

print(paste0("Cumulative probability. P(X>x): ", round(prob.cumulative.gt, 4)))
```

```
  if (is_normal_approx) {
    print(paste0("    Normal approximation. P(X>x): ",
                 round(prob.cumulative.gt.approx, 4)))
    print(paste0("    Normal approximation w/ continuity correction. P(X>x): ",
                 round(prob.cumulative.gt.approx.corr, 4)))
  }

    print(paste0("Cumulative probability. P(X>=x): ", round(prob.cumulative.ge, 4)))
  if (is_normal_approx) {
    print(paste0("    Normal approximation. P(X>=x): ",
                 round(prob.cumulative.ge.approx, 4)))
    print(paste0("    Normal approximation w/ continuity correction. P(X>=x): ",
                 round(prob.cumulative.ge.approx.corr, 4)))
  }
}
```

I ran it for this problem (incorrectly).

```
binom_summary(15, 100, 0.5)
```

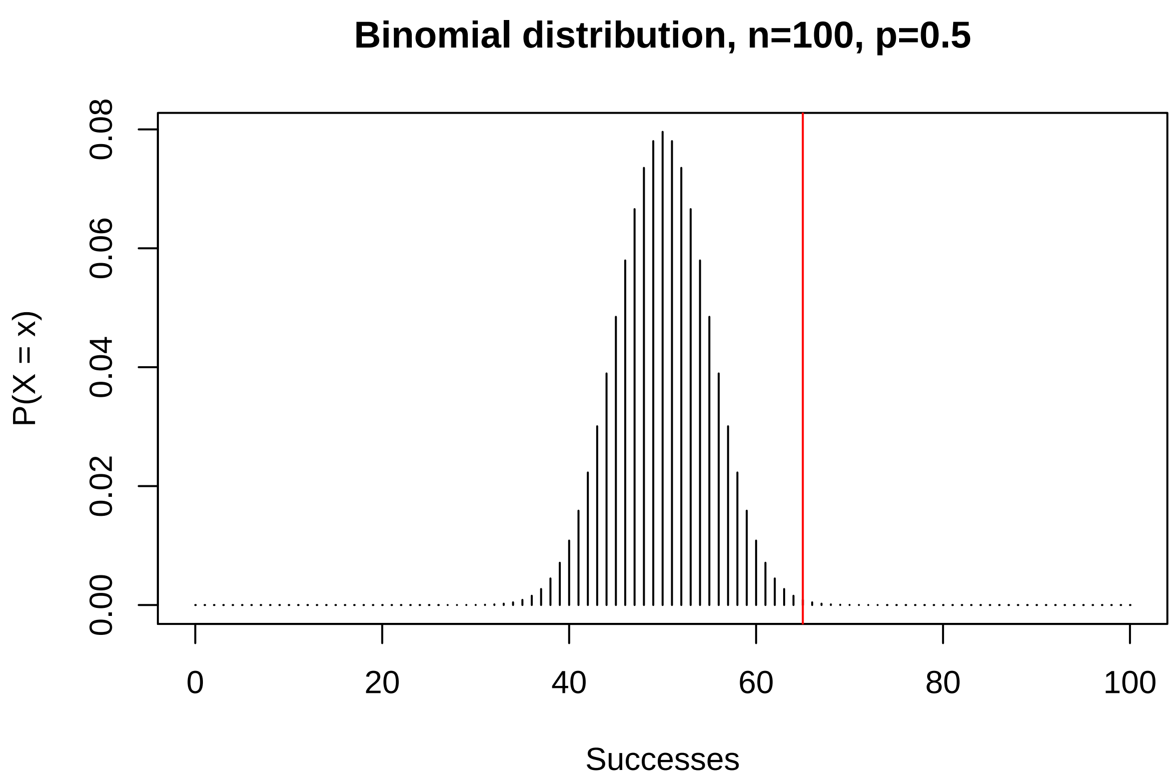



```
## [1] "mu: 50"
## [1] "sigma: 5"
## [1] "Trials: 100"
## [1] "Successes: 15"
## [1] "Binomial probability. P(X=x): 0"
```

```
## [1] "Cumulative probability. P(X<x): 0"
## [1] "    Normal approximation. P(X<x): 0"
## [1] "    Normal approximation w/ continuity correction. P(<x): 0"
## [1] "Cumulative probability. P(X<=x): 0"
## [1] "    Normal approximation. P(X<=x): 0"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 0"
## [1] "Cumulative probability. P(X>x): 1"
## [1] "    Normal approximation. P(X>x): 1"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 1"
## [1] "Cumulative probability. P(X>=x): 1"
## [1] "    Normal approximation. P(X>=x): 1"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 1"
```

How great are pictures?! The successes I input, 15, is plotted with the red line. Instantly, I see my mistake. The successes I'm supposed to check don't amount to 15, they are the mean PLUS 3 standard deviations, and MINUS, too. Let's try again.

```
binom_summary(50 + 15, 100, 0.5)
```




```
## [1] "mu: 50"
## [1] "sigma: 5"
## [1] "Trials: 100"
## [1] "Successes: 65"
## [1] "Binomial probability. P(X=x): 0.0009"
## [1] "Cumulative probability. P(X<x): 0.9982"
```

```
## [1] "    Normal approximation. P(X<x): 0.9987"
## [1] "    Normal approximation w/ continuity correction. P(<x): 0.9981"
## [1] "Cumulative probability. P(X<=x): 0.9991"
## [1] "    Normal approximation. P(X<=x): 0.9987"
## [1] "    Normal approximation w/ continuity correction. P(<=x): 0.999"
## [1] "Cumulative probability. P(X>x): 0.0009"
## [1] "    Normal approximation. P(X>x): 0.0013"
## [1] "    Normal approximation w/ continuity correction. P(X>x): 0.001"
## [1] "Cumulative probability. P(X>=x): 0.0018"
## [1] "    Normal approximation. P(X>=x): 0.0013"
## [1] "    Normal approximation w/ continuity correction. P(X>=x): 0.0019"
```

We want both tails, so twice that tail probability is 0.0036, rounded to 4 places. I'll calculate that again without rounding. The answer, 0.0035176. That is consistent with my answer for the previous problem which said that the probability *at most* is 0.1111111. There can be a big margin around the results we calculate using Chebyshev's Inequality.