Central limit theorem (中央極限定理)

- Central limit theorem. Suppose that $X_1, \ldots, X_n$ are IID with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Let $\bar{X} = (X_1 + \cdots + X_n)/n$. Then the distribution of $\dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$ is approximately $N(0, 1)$ in the sense that

$$P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq x\right) \approx P(N(0, 1) \leq x)$$

for every x for large n.

  - Recall that from (6) and (7) in the handout "Mean, variance and standard deviation", we have

$$E(\bar{X}) = \mu \text{ and } Var(\bar{X}) = \frac{\sigma^2}{n},$$

so

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{\bar{X} - E(\bar{X})}{(Var(\bar{X}))^{1/2}}$$

is a random variable with mean 0 and variance 1.

- Raw normal approximation to $Bin(n, p)$. Suppose that $X_1, \cdots, X_n$ are IID $Bin(1, p)$, then $E(X_1) = p$ and $Var(X_1) = p(1 - p)$. Let

$$Z = \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1 - p)}},$$

then the distribution of Z is approximately $N(0, 1)$ for large n. Express $n\bar{X}$ in terms of Z and we have $n\bar{X} = np + \sqrt{np(1 - p)}Z$, so

$$\begin{aligned} P(Bin(n, p) \leq x) &= P(np + \sqrt{np(1 - p)}Z \leq x) \\ &= P\left(Z \leq \frac{x - np}{\sqrt{np(1 - p)}}\right) \\ &\approx P\left(N(0, 1) \leq \frac{x - np}{\sqrt{np(1 - p)}}\right) \\ &= P(N(np, np(1 - p)) \leq x). \end{aligned}$$

The formula

$$P(Bin(n, p) \leq x) \approx P(N(np, np(1 - p)) \leq x) \tag{1}$$

provides one way to approximate $P(Bin(n, p) \leq x)$.

- Normal approximation to $Bin(n,p)$ with correction for continuity. Suppose that $X_1, \cdots, X_n$ are IID $Bin(1,p)$, then $E(X_1) = p$ and $Var(X_1) = p(1-p)$.

  When $x = k$ is an integer, the formula

  $$P(Bin(n,p) \le k) \approx P(N(np, np(1-p)) < k + 0.5), \qquad (2)$$

  provides a better approximation than the formula in (1).

  Example 1. Suppose that $X \sim Bin(80, 0.7)$. Approximate $P(X < 60)$ and $P(X \ge 60)$ using normal probabilities.

  Sol. $E(X) = 80 \times 0.7 = 56$ and $Var(X) = 80 \times 0.7 \times (1 - 0.7) = 16.8$.

  $$\begin{aligned} P(X < 60) &= P(X \le 59) \\ &\approx P(N(56, 16.8) < 59 + 0.5) \\ &= P\left(N(0,1) < \frac{59.5 - 56}{\sqrt{16.8}}\right) \\ &\approx P(N(0,1) < 0.85) \approx 0.5 + 0.3023 = 0.8023. \end{aligned}$$

  $P(X \ge 60) = 1 - P(X < 60) \approx 1 - 0.8023 = 0.1977$.

- The probabilities $P(Bin(n,p) \le x)$ and $P(N(\mu, s^2) \le x)$ can be computed using R commands `pbinom` and `pnorm` respectively:

  - $P(Bin(n,p) \le x)$ is computed using `pbinom(`x`, size=`n`, prob=`p`)`
  - $P(N(\mu, s^2) \le x)$ is computed using `pnorm(`x`, mean=`μ`, sd=`s`)`

- Compare $P(Bin(n,p) \le x)$ with `pbinom(`x`, size=`n`, prob=`p`)` when $p = 0.5$, $n = 10$.

```
binom.cdf.values <- function(n,p){
 ans <- rep(0, n+1)
 ans[1] <- (1-p)^n
 for (i in 2:(n+1)){ ans[i] <- ans[i-1] + choose(n,i-1)*p^(i-1)*(1-p)^(n-i+1) }
 return(ans)
 }
 binom.cdf <- function(x){ pbinom(x, size=10, prob=0.5) }
 curve(binom.cdf, 0, 10, n=1000)
 points((0:10), binom.cdf.values(10, 0.5), col=2)
```

  The output graph shows that `pbinom(`x`, size=10, prob=0.5)` agrees with $P(Bin(10, 0.5) \le x)$ for $x = 0, 1, \ldots, 10$.

- Compare $P(N(0,1) \le x)$ with `pnorm(`x`, mean=0, sd=1)` for $x \in [-3, 3]$.

```
density.fun <- function(x){ exp(-x^2/2)/sqrt(2*pi) }
cdf.fun <- function(x){ integrate(density.fun, -Inf, x)$value }
x <- ((-300):300)/100
y <- x
for (i in 1:length(x)){ y[i] <- cdf.fun(x[i]) }
plot(x,y,type="l")

f <- function(x){ pnorm(x, mean=0, sd=1) }
curve(f,-3,3, add=T, col=2)
```

The output graph shows that $P(N(0,1) \leq x)$ agrees with `pnorm(`x`, mean=0, sd=1)` for $x \in [-3,3]$.

- An experiment. We plot $P(N(np, np(1-p) \leq x)$ and $P(Bin(n,p) \leq x)$ for 1000 x's in $[-1,40]$ with $(n,p) = (40, 0.5)$. We also plot the probabilities based on normal approximation with continuity correction. The R scripts and the outputs are given below.

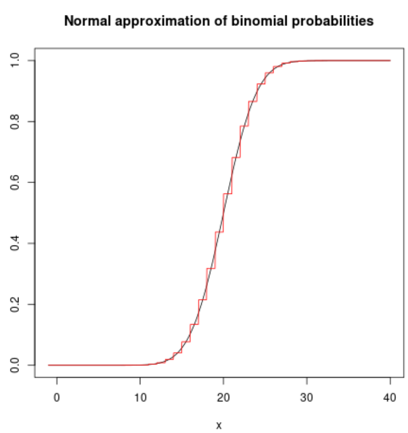
```
x <- seq(-1,40, length=1000)
n <- 40; p <- 0.5

#plot the cdf of N(n*p, n*p*(1-p))
plot(x, pnorm(x, mean=n*p, sd=sqrt(n*p*(1-p))), type="l", ylab="")

#add the cdf of Bin(n,p)
lines(x, pbinom(x, size=n, prob=p), col="red")

title("Normal approximation of binomial probabilities")
```

The plot is shown below. $P(Bin(n,p) \leq x)$'s are plotted in red. The plot shows that the normal probabilities approximate the binomial probabilities well.



We can also add the approximation with continuity correction:

```
x1 <- 0:40
points(x1, pnorm(x1+0.5, mean=n*p, sd=sqrt(n*p*(1-p))), col="blue")
```

The plot is shown below. The approximation probabilities with continuity correction are plotted in blue. The plot shows that the binomial probabilities can be better approximated by the normal probabilities with continuity correction.