QCB 408 / 508 – Notes on Week 3

Zihong Chen

2/26/2020

Summary

- Probabilistic models of RNA-seq data
- Negative binomial model: "2 steps" of Poisson sampling process
- Generalization: Gamma - Poisson distribution
- Summary of parameters and random variables in RNA-seq model
- Sums of random variables
- Convergence of random variables
- Convergence in distribution
- Convergence in probability
- Almost sure convergence
- Strong law of large numbers
- Central limit theorem

Probabilistic models of RNA-seq data

In most cases, RNA-seq data are used to compare different gene expression patterns between 2 or more biological conditions (e.g., healthy versus disease). But for modeling, we start from a simplier situation, that all the RNA-seq data are from a single biological condition or population.

Assuming:

- $i = 1, 2, ..., m$ as the number of genes
- $j = 1, 2, ..., n$ as the number of obsevations
- Y_{ij} as the read counts we observed from RNA-seq experiments, for gene i in observation j

Given Y_{ij}, our target is to find the true proportion of gene i expression (or relative expression), a_i. Note that most a_i are small, and $\sum a_i = 1$.

Negative binomial model: "2 steps" of Poisson sampling process

As an instrutive, idealized example to follow, RNA-seq data can be generated in 2 steps:

- Step 1: Sample cells and mRNA molecules
- Step 2: Sequence mRNA molecules and obtain counts

The general idea for the first model is to consider these 2 steps as 2 Poisson process.

Step 1: Sample cells and mRNA molecules

For step 1, we can define random raviables:

- M_j: the number of mRNA molecules sampled in observation j
- X_{ij}: the number of mRNA molecules sampled for gene i in observation j

Assuming completely random sampling of mRNA molecules:

$$X_{ij}|M_j \sim \text{Binomial}(M_j, a_i)$$

Given that M_j is very large and a_i is very small, Binomial distribution becomes Poisson distribution:

$$Var(X_{ij}|M_j) = M_j a_i (1 - a_i) \approx M_j a_i = E(X_{ij}|M_j)$$

$$X_{ij}|M_j \sim \text{Poisson}(M_j a_i)$$

Then we can further define Π_{ij} as the random proportion of gene i mRNA in observation j:

$$\Pi_{ij} = \frac{X_{ij}}{M_j}$$

Its expected value and variance are:

$$\begin{aligned}
E[\Pi_{ij}] &= E[E[\Pi_{ij}|M_j]] \\
&= E[E[\frac{X_{ij}}{M_j}|M_j]] \\
&= E[\frac{M_j a_i}{M_j}|M_j] \\
&= E[a_i|M_j] \\
&= a_i
\end{aligned}$$

$$\begin{aligned}
Var(\Pi_{ij}) &= E[Var(\Pi_{ij}|M_j)] + Var(E[\Pi_{ij}|M_j]) \\
&= E[Var(\frac{X_{ij}}{M_j}|M_j)] + Var(a_i|M_j) \\
&= E[\frac{M_j a_i}{M_j^2}|M_j] + 0 \\
&= a_i E[\frac{1}{M_j}]
\end{aligned}$$

Note that $Var(\Pi_{ij})$ here represents biological variance, which is the same for different observation j.

Step 2: Sequence mRNA molecules and obtain counts

For step 2, we can define:

- D_j: the "read depth" of observation j, which is the total number of reads from observation j
- d_j: the observed "read depth" of observation j

Note that here d_j is the actual observation, while D_j is a random variable that models the obtain data. When used to build conditional probability, we say d_j in the condition, which means $D_j = d_j$.

Similarly, assuming completely random sampling, we have:

$$\begin{aligned}
Y_{ij}|\Pi_{ij}, d_j &\sim \text{Binomial}(d_j, \Pi_{ij}) \\
&\sim \text{Poisson}(d_j \Pi_{ij})
\end{aligned}$$

Therefore, the expected value and variance of Y_{ij} are:

$$\begin{aligned}
E[Y_{ij}] &= E[E[Y_{ij}|\Pi_{ij}, d_j]] \\
&= E[d_j \Pi_{ij}] \\
&= d_j E[\Pi_{ij}] \\
&= d_j a_i
\end{aligned}$$

$$\begin{aligned}
Var(Y_{ij}) &= E[Var(Y_{ij}|\Pi_{ij}, d_j)] + Var(E[Y_{ij}|\Pi_{ij}, d_j]) \\
&= E[d_j \Pi_{ij}] + Var(d_j \Pi_{ij}) \\
&= d_j E[\Pi_{ij}] + d_j^2 Var(\Pi_{ij}) \\
&= d_j a_i + d_j^2 a_i E[\frac{1}{M_j}] \\
&> E[Y_{ij}]
\end{aligned}$$

Note that $Y_{ij}|\Pi_{ij}, d_j$ is Poisson, while Y_{ij} becomes over-dispersed Poisson, which is true for RNA-seq data which usually have larger variance than mean.

Estimation of Π_{ij} and a_i

Now that we have built up the model, we can try to estimate Π_{ij}, which is the random proportion of gene i mRNA in observation j, and a_i, which is the the true proportion of gene i expression (or the population mean). For $\hat{\Pi}_{ij}$:

$$\hat{\Pi}_{ij} = \frac{Y_{ij}}{d_j}$$

$$\begin{aligned} E[\hat{\Pi}_{ij}] &= E[\frac{Y_{ij}}{d_j}] \\ &= \frac{d_j a_i}{d_j} \\ &= a_i \end{aligned}$$

$$\begin{aligned} Var(\hat{\Pi}_{ij}) &= Var(\frac{Y_{ij}}{d_j}) \\ &= \frac{Var(Y_{ij})}{d_j^2} \\ &= \frac{a_i}{d_j} + a_i E[\frac{1}{M_j}] \end{aligned}$$

Note that $Var(\Pi_{ij}) = a_i E[\frac{1}{M_j}]$ here accounts for the biological variance, while the rest $\frac{a_i}{d_j}$ is the technical variance, which goes to 0 as $d_j \to \infty$.

For $\hat{a}_i$:

$$\hat{a}_i = \frac{\sum_{j=1}^n \hat{\Pi}_{ij}}{n}$$

$$\begin{aligned} E[\hat{a}_i] &= \frac{E[\sum_{j=1}^n \hat{\Pi}_{ij}]}{n} \\ &= \frac{\sum_{j=1}^n E[\hat{\Pi}_{ij}]}{n} \\ &= a_i \end{aligned}$$

$$\begin{aligned} Var(\hat{a}_i) &= \frac{Var(\sum_{j=1}^n \hat{\Pi}_{ij})}{n^2} \\ &= \frac{\sum_{j=1}^n Var(\hat{\Pi}_{ij})}{n^2} \\ &= \frac{\sum_{j=1}^n \frac{Var(Y_{ij})}{d_j^2}}{n^2} \\ &= \frac{a_i}{n^2} \sum_{j=1}^n \frac{1}{d_j} + \frac{a_i}{n^2} \sum_{j=1}^n E[\frac{1}{M_j}] \end{aligned}$$

Here $Var(\hat{a}_i)$ is still split into the technical and biological parts.

Since all the d_j's are the data obtained, the key is to solve the rest part $E[\frac{1}{M_j}]$, which comes from $Var(Y_{ij})$, or the distribution of Y_{ij}.

To simplify, we assume that M_j are i.i.d., then $E[\frac{1}{M_j}]$ is the same for all j. Then we can introduce coefficient of variation CV (biological) to rewrite $Var(Y_{ij})$:

$$\begin{aligned}
CV &= \frac{\sqrt{Var(\Pi_{ij})}}{a_i} \\
CV^2 &= \frac{Var(\Pi_{ij})}{a_i^2} = \frac{1}{a_i} E[\frac{1}{M_j}] \equiv \phi_i \\
Var(Y_{ij}) &= d_j a_i + d_j^2 a_i E[\frac{1}{M_j}] \\
&= d_j a_i + (d_j a_i)^2 \frac{1}{a_i} E[\frac{1}{M_j}] \\
&= \mu_{ij} + \mu_{ij}^2 \phi_i
\end{aligned}$$

where $\mu_{ij} = d_j a_i$ (observed read depth times population mean). It indicates that Y_{ij} follows the negative binomial distribution, where ϕ_i can be estimated by "borrowing strength" across genes with similar ϕ_i.

Negative binomial distribution

Considering Bernoulli trials with $Pr(success) = p$, Y is the number of failures before the r^{th} success, then Y follows the negative binomial distribution with parameter r and p:

$$Y \sim \text{NegBin}(r, p)$$

Its pmf is similar to binomial distribution, except that the last trail is fixed as "success":

$$Pr(Y = y) = \binom{r+y-1}{y} p^r (1-p)^y, y = 0, 1, 2, ...$$

For its expected value and variance, we can calculate $E[Y^n]$ first:

$$\begin{aligned}
E[Y^n] &= \sum_{y=0}^{\infty} y^n \binom{r+y-1}{y} p^r (1-p)^y \\
&= 0 + \sum_{y=1}^{\infty} y^n \frac{(r+y-1)!}{(r-1)!y!} p^r (1-p)^y \\
&= \frac{r(1-p)}{p} \sum_{y=1}^{\infty} y^{n-1} \frac{(r+y-1)!}{r!(y-1)!} p^{r+1} (1-p)^{y-1} \\
&= \frac{r(1-p)}{p} E[(Z+1)^{n-1}]
\end{aligned}$$

where $Z \sim \text{NegBin}(r+1, p)$. Therefore:

$$\begin{aligned}
n = 1, E[Y] &= \frac{r(1-p)}{p} E[(Z+1)^0] = \frac{r(1-p)}{p} \\
n = 2, E[Y^2] &= \frac{r(1-p)}{p} E[(Z+1)^1] \\
&= \frac{r(1-p)}{p}(1 + E[Z]) \\
&= \frac{r(1-p)}{p}[1 + \frac{(r+1)(1-p)}{p}] \\
&= \frac{r^2(1-p)^2 + r(1-p)}{p^2} \\
Var(Y) &= E[Y^2] - E[Y]^2 \\
&= \frac{r^2(1-p)^2 + r(1-p)}{p^2} - \frac{r^2(1-p)^2}{p^2} \\
&= \frac{r(1-p)}{p^2} \\
&= \frac{r(1-p)}{p} + [\frac{r(1-p)}{p}]^2 \frac{1}{r}
\end{aligned}$$

RNA-seq data under the above model is therefore sometimes model as a $Y_{ij} \sim NegBin(r_i, p_{ij})$, which can be reparameterized by $\mu_{ij} = \frac{r_i(1-p_{ij})}{p_{ij}}, \phi_i = \frac{1}{r_i}$.

Generalization: Gamma - Poisson distribution

In the previous model, we modeled Y_{ij} as 2 steps of Poisson process, which gives a negative binomial:

$$Y_{ij}|\Pi_{ij}, d_j \sim \text{Poisson}(d_j \Pi_{ij})$$

$$\Pi_{ij} = \frac{X_{ij}}{M_j}, X_{ij}|M_j \sim \text{Poisson}(M_j a_i)$$

While in our definition of negative binomial distribution, r has to be a positive integer. To generalize r to all real numbers, mathematically Y_{ij} becomes marginally a Gamma - Poisson random variable. It means that in the model, we keep the Poisson in $Y_{ij}|\Pi_{ij}, d_j$ while assume that $\lambda_{ij} = d_j \Pi_{ij}$ follows gamma distribution:

$$Y_{ij}|\lambda_{ij} \sim \text{Poisson}(\lambda_{ij})$$
$$\lambda_{ij} \sim \text{Gamma}(\alpha_{ij}, \beta_{ij})$$

Several facts on Gamma distribution:

$$f(\lambda; \alpha, \beta) = \frac{\lambda^{\beta-1} e^{-\lambda/\alpha}}{\alpha^\beta \Gamma(\beta)}$$

$$E[\lambda] = \alpha\beta, Var(\lambda) = \alpha^2\beta$$

And several facts on Gamma - Poisson distribution:

$$f(y; \alpha, \beta) = \frac{\Gamma(y+\beta)}{\Gamma(\beta) y!} (\frac{1}{1+\alpha})^\beta (\frac{\alpha}{1+\alpha})^y$$

$$E[Y] = \alpha\beta, Var(Y) = \alpha\beta + \alpha^2\beta$$

Note that when β is an positive integer, the pdf of Gamma - Poisson is the same as negative binomial:

$$\Gamma(y+\beta) = (y+\beta-1)!, \Gamma(\beta) = (\beta-1)!$$

$$f(y;\alpha,\beta) = \binom{y+\beta-1}{y}(\frac{1}{1+\alpha})^{\beta}(\frac{\alpha}{1+\alpha})^{y}$$

$$\alpha = \frac{1-p}{p}, \beta = r$$

Conbine with the reparameterization of the negative binomial, we have:

$$\phi_i = \frac{1}{r_i} = \frac{1}{a_i}E[\frac{1}{M_j}], \mu_{ij} = \frac{r_i(1-p_{ij})}{p_{ij}} = d_j a_i$$

$$r_i = a_i E^{-1}[\frac{1}{M_j}], p_{ij} = \frac{1}{1+d_j E[\frac{1}{M_j}]}$$

$$\beta_{ij} = a_i E^{-1}[\frac{1}{M_j}], \alpha_{ij} = d_j E[\frac{1}{M_j}]$$

Mean-variance relationship

The negative binomial distribution and the Gamma - Possion distribution are similar because of the same mean - variance relationship.

Suppose $\lambda > 0$ is a random variable with $E(\lambda)$ and $Var(\lambda)$, and

$$Y|\lambda \sim \text{Poisson}(\lambda)$$

$$\begin{aligned} Var(Y) &= E[Var(Y|\lambda)] + Var(E[Y|\lambda]) \\ &= E[\lambda] + Var(\lambda) \end{aligned}$$

As long as λ is not symmetric, $E[\lambda]$ will always appear in $Var(\lambda)$, which defines the mean - variance relationship. For example, the Poisson distribution has a linear relationship $E[\lambda] = Var(\lambda)$, while the Normal distribution is symmetric and thus the mean and variance are independent parameters.

Both the negative binomial and the Gamma - Poisson has a quadratic mean - variance relationship on $\lambda_{ij}(d_j\Pi_{ij})$:

$$E[d_j\Pi_{ij}] = d_j a_i, Var[d_j\Pi_{ij}] = d_j^2 a_i E[\frac{1}{M_j}] = E[d_j\Pi_{ij}]^2\phi_i$$

$$E[\lambda_{ij}] = \alpha_{ij}\beta_{ij}, Var(\lambda_{ij}) = \alpha_{ij}^2\beta_{ij} = E[\lambda_{ij}]^2\phi_i$$

Summary of parameters and random variables in RNA-seq model

`Table 1 Summary of parameters in RNA-seq models`

Parameters	Definition
i	numbers of genes
j	numbers of observations
a_i	the true proportion of gene i expression

Parameters	Definition
$r_i = a_i E^{-1}[1/M_j]$	parameters for negative binomial distribution
$p_{ij} = 1/(1 + d_j E[1/M_j])$	parameters for negative binomial distribution
$\alpha_{ij} = d_j E[1/M_j]$	parameters for Gamma - Poisson distribution
$\beta_{ij} = a_i E^{-1}[1/M_j]$	parameters for Gamma - Poisson distribution
$CV = \sqrt{Var(\Pi_{ij})}/a_i$	coefficient of (biological) variance
$\mu_{ij} = d_j a_i$	parameters for both negative binomial and Gamma - Poisson
$\phi_i = E[1/M_j]/a_i$	parameters for both negative binomial and Gamma - Poisson

Table 2 Summary of random variables in RNA-seq models

Random variables	Definition	E[x]	Var(x)
Y_{ij}	read counts for gene i in observation j	$d_j a_i$	$d_j a_i + d_j^2 a_i E[\frac{1}{M_j}]$
M_j	the number of mRNA molecules sampled in observation j		
X_{ij}	the number of mRNA molecules sampled for gene i in observation j		
$\Pi_{ij} = X_{ij}/M_j$	the random proportion of gene i mRNA in observation j	a_i	$a_i E[\frac{1}{M_j}]$
$D_j(d_j)$	the total number of reads in observation j		

Random variables	Definition	E[x]	Var(x)
$\hat{\Pi}_{ij} = Y_{ij}/d_j$	estimator of Π_{ij}	a_i	$\frac{a_i}{d_j} + a_i E[\frac{1}{M_j}]$
$\hat{a}_i = (\sum_{j=1}^{n} \hat{\Pi}_{ij})/n$	estimator of a_i	a_i	$\frac{a_i}{n^2} \sum_{j=1}^{n} \frac{1}{d_j} + \frac{a_i}{n^2} \sum_{j=1}^{n} E[\frac{1}{M_j}]$
$\lambda_{ij} = d_j \Pi_{ij}$		$d_j a_i$	$d_j^2 a_i E[\frac{1}{M_j}]$

Sums of random variables

If X is a random variable and a, b are constants, then:

$$E[a + bX] = a + bE[X]$$

$$Var(a + bX) = b^2 Var(X)$$

Let $X_1, X_2, ..., X_n$ be n random variables, then:

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$$

$$\begin{aligned} Var(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j) \end{aligned}$$

Suppose $X_1, X_2, ..., X_n$ are independent:

$$\sum_{i \neq j} Cov(X_i, X_j) = 0$$

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

In this case, we further define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, then:

$$\begin{aligned} E[\bar{X}_n] &= E[\frac{1}{n} \sum_{i=1}^{n} X_i] \\ &= \frac{1}{n} E[\sum_{i=1}^{n} X_i] \\ &= \frac{1}{n} \sum_{i=1}^{n} E[X_i] \end{aligned}$$

$$\begin{aligned} Var(\bar{X}_n) &= Var(\frac{1}{n} \sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) \end{aligned}$$

When $E[X_1] = E[X_2] = ... = E[X_n] = \theta$, then $E[\bar{X}_n] = \theta$.

When $Var(X_1) = Var(X_2) = ... = Var(X_n) = \tau^2$, then $Var(\bar{X}_n) = \frac{\tau^2}{n}$.

Example:

$$X_1, X_2, ..., X_n \overset{i.i.d}{\sim} \text{Normal}(\mu, \sigma^2)$$

$$X_1 + X_2 + ... + X_n \sim \text{Normal}(n\mu, n\sigma^2)$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \text{Normal}(\mu, \frac{\sigma^2}{n})$$

$$E[\bar{X}_n] = \mu = E[X_i], i = 1, 2, ..., n$$

$$Var(\bar{X}_n) = \frac{\sigma^2}{n} = \frac{Var(X_i)}{n}, i = 1, 2, ..., n$$

Convergence of random variables

Let $Z_1, Z_2, ... = \{Z_i\}$ be a sequence of random variables, for example:

$$Z_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\text{or} \quad Z_n \sim \text{Binomial}(n, p)$$

It is useful to be able to determine a limiting value or distribution of $\{Z_i\}$.

Convergence in distribution

$\{Z_i\}$ converges in distribution to a random variable W, written

$$Z_n \xrightarrow{D} W$$

if

$$F_{Z_n}(y) = Pr(Z_n \leq y) \to Pr(W \leq y) = F_Z(y)$$

for all $y \in \mathbb{R}$, as $n \to \infty$.

Convergence in distribution is the weakest form of convergence typically discussed, since it is implied by all other types of convergence mentioned here. However, convergence in distribution is very frequently used in practice; most often it arises from application of the **central limit theorem** (Wikipedia).

Convergence in probability

$\{Z_i\}$ converges in probability to a random variable W, written

$$Z_n \xrightarrow{P} W$$

if

$$Pr(|Z_n - W| \leq \epsilon) \to 1$$

for all $\epsilon > 0$, as $n \to \infty$. Note that when W is not a random variable but a fixed value, we can still have such convergence.

Convergence in probability implies convergence in distribution.In the opposite direction, convergence in distribution implies convergence in probability when the limiting random variable W is a constant (Wikipedia).

Almost sure convergence

$\{Z_i\}$ converges almost surely (or "almost everywhere", or "with probability 1") to a random variable W, written

$$Z_n \xrightarrow{a.s.} W$$

if

$$Pr\left(\left\{w : |Z_n(w) - W(w)| \xrightarrow{n\to\infty} 0\right\}\right) = 1$$

It may also be the case that W is a constant rather than a random variable.

Almost sure convergence implies the above 2 types of convergence. It is the notion of convergence used in the **strong law of large numbers** (Wikipedia).

Strong law of large numbers

The law of large numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times.

Suppose $X_1, X_2, ..., X_n$ are n i.i.d random variables with population mean $E[X_i] = \mu$ where $E[|X_i|] < \infty$, then as $n \to \infty$:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu$$

According to the law, the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer as more trials are performed (Wikipedia).

Central limit theorem

Suppose $X_1, X_2, ..., X_n$ are n i.i.d random variables with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$, then as $n \to \infty$,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \text{Normal}(0, \sigma^2)$$

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} \text{Normal}(0, 1)$$

The central limit theorem establishes that, sample mean $\bar{X}_n$ tends toward a normal distribution, even if the original variables themselves $(X_1, X_2, ..., X_n)$ are not normally distributed.It implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions (Wikipedia).

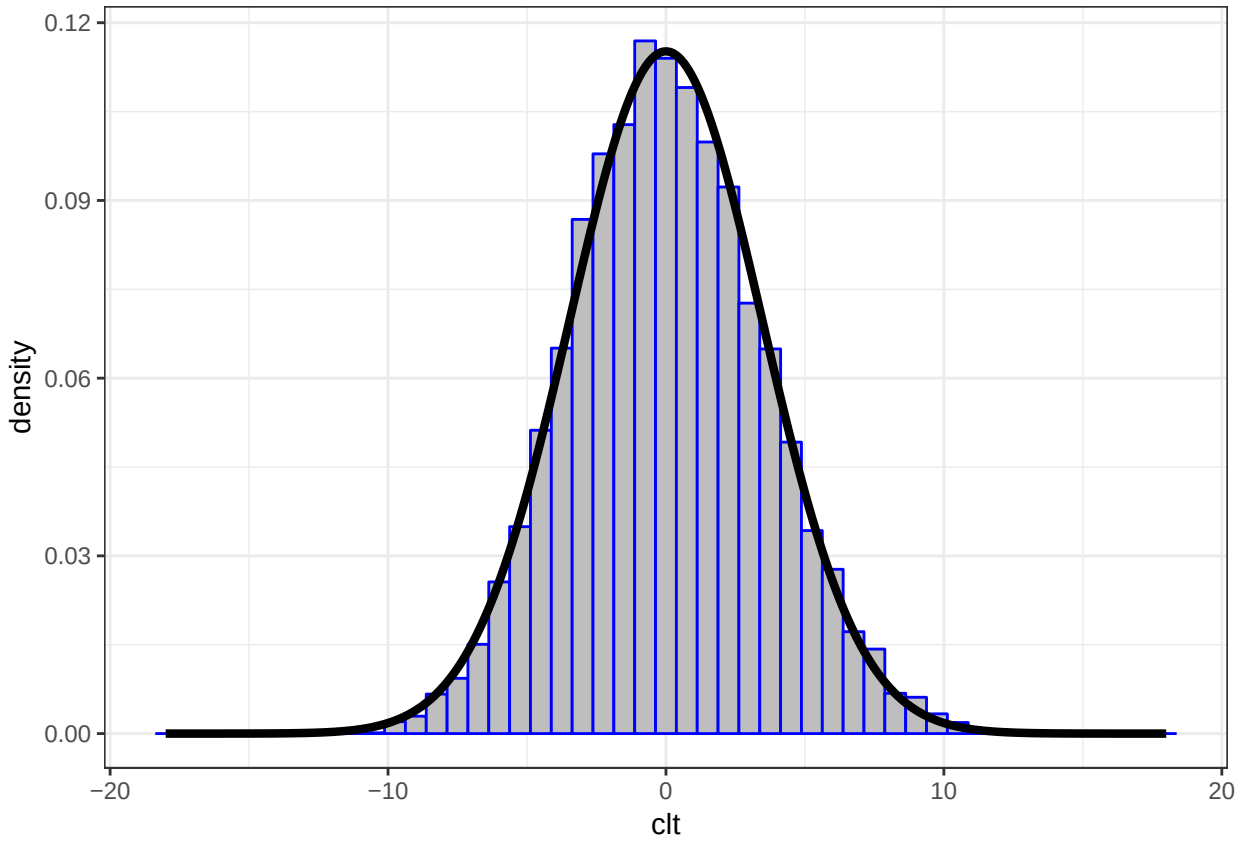
Example:

In our textbook (FAS), we have an example of Poisson. Let's try another distribution here.

Suppose $X_1, X_2, ..., X_{50}$ as an iid sequence of Gamma$(3, 2)$, which gives $E[\bar{X}_n] = 6$ and $Var(\bar{X}_n) = 12$.

We then form $\sqrt{50}(\bar{X}_n - 6)$ over 10,000 times and compare their distribution to Normal$(0, 12)$:

```
> x <- replicate(n=1e4, expr=rgamma(n=50, shape=3, scale=2), simplify="matrix");
> x_bar <- apply(x, 2, mean);
> clt <- sqrt(50)*(x_bar - 6);
> df <- data.frame(clt=clt, x=seq(-18,18,length.out=1e4),
+                  y=dnorm(seq(-18,18,length.out=1e4), sd=sqrt(12)))
> ggplot(data=df) +
+    geom_histogram(aes(x=clt, y=..density..), color="blue", fill="gray", binwidth=0.75) +
+    geom_line(aes(x=x, y=y), size=1.5)
```



Session Information

```
> sessionInfo()
R version 3.6.0 (2019-04-26)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS  10.15.3

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] gapminder_0.3.0 forcats_0.4.0   stringr_1.4.0   dplyr_0.8.1
 [5] purrr_0.3.2     readr_1.3.1     tidyr_0.8.3     tibble_2.1.1
 [9] ggplot2_3.1.1   tidyverse_1.2.1 knitr_1.22

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.1       cellranger_1.1.0 pillar_1.4.0     compiler_3.6.0
 [5] plyr_1.8.4       tools_3.6.0      digest_0.6.18    lubridate_1.7.4
 [9] jsonlite_1.6     evaluate_0.13    nlme_3.1-140     gtable_0.3.0
```

```
[13] lattice_0.20-38  pkgconfig_2.0.2  rlang_0.3.4       cli_1.1.0
[17] rstudioapi_0.10  yaml_2.2.0       haven_2.1.0       xfun_0.7
[21] withr_2.1.2      xml2_1.2.0       httr_1.4.0        hms_0.4.2
[25] generics_0.0.2   grid_3.6.0       tidyselect_0.2.5  glue_1.3.1
[29] R6_2.4.0         readxl_1.3.1     rmarkdown_1.12    modelr_0.1.4
[33] magrittr_1.5     backports_1.1.4  scales_1.0.0      htmltools_0.3.6
[37] rvest_0.3.3      assertthat_0.2.1 colorspace_1.4-1  labeling_0.3
[41] stringi_1.4.3    lazyeval_0.2.2   munsell_0.5.0     broom_0.5.2
[45] crayon_1.3.4
```