Linear Algebra Study group

Bill Chung

January 14, 2022

```
library(far)
library(MASS)
library(pracma)
```

Terms:

- size of vector, span
- linear combination
- subspace: contains zero, closed under addition and multiplcation
- norm and dot product, unit vector
- dependent and independent vectors
- rref, C(A)
- Singular (degenerate) and nonsingular matrix
- Covariance matrix
- $\mathbb{I}$

Law of Total Probability

$$\begin{aligned} P(B) &= \sum_i P(B \cap A_i) \\ &= \sum_i P(B|A_i)P(A_i) \end{aligned}$$

$$\begin{aligned} P(B) &= \sum_i P(B \cap A_i) \\ &= \sum_i P(B|A_i)P(A_i) \end{aligned}$$

Conditional probability

$$\begin{aligned} &\text{Conditional probability} \\ f_{Y|X}(y|x) &= Pr[Y = y|X = x] \\ &= \frac{f(x,y)}{f_X(x)} \end{aligned}$$

Bayes' Rule and Conditional probability

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \end{aligned}$$

Joint probability

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ f(x, y) &= Pr[X = x, Y = y] \;\; \forall x, y \in R \\ F(x, y) &= Pr[X \leq x, Y \leq y] \;\; \forall x, y \in R \end{aligned}$$

Marginal probability

$$\begin{aligned} f_Y &= Pr[Y = y] \\ &= \sum_{x \in Supp[X]} f(x, y), \forall y \in R \end{aligned}$$

Law of Iterated Expectation

$$E[Y] = E[E[Y|X]]$$

Law of Total Variance

$$V[Y] = E[V[Y|X]] + V[E[Y|X]]$$

Operators

Expectation Operator

$$\begin{aligned} E[x] &= \sum_x x f(x) \\ &= \int_{-\infty}^{\infty} x f(x) dx \end{aligned}$$

$$\begin{aligned} E[g(x)] &= \sum_x g(x) f(x) \\ &= \int_{-\infty}^{\infty} g(x) f(x) dx \end{aligned}$$

$$E[\vec{x}] = (E[x_1], E[x_2] ... E[x_n])$$

$$\begin{aligned} E[h(X, Y)] &= \sum_x \sum_y h(x, y) f(x, y) \\ &= \int \int h(x, y) f(x, y) dy dx \end{aligned}$$

Conditional Expectation

$$\begin{aligned} &\text{Conditional Expectation} \\ E[Y|X=x] &= \sum_y y f_{Y|X}(y|x) \quad \forall x \in Supp[X] \\ &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \quad \forall x \in Supp[X] \end{aligned}$$

$$\begin{aligned} E[h(X,Y)|X=x] &= \sum_y h(x,y) f_{Y|X}(y|x) \quad \forall x \in Supp[X] \\ &= \int_{-\infty}^{\infty} h(x,y) f_{Y|X}(y|x) dy \quad \forall x \in Supp[X] \end{aligned}$$

Variance

$$V[X] = \sigma_X^2 = \sum_{i=1}^{N} (X_i - E(X))^2 p_i$$

Sample variance

- Population variance estimated from sample

$$\hat{V}(X) = S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$\hat{V}(X) = \frac{n}{n-1}(\bar{X^2} - (\bar{X})^2)$$

Properties of variance

$$\begin{aligned} V[X] &= E[(X - E[X])^2] \\ &= E[X^2] - E[X]^2 \end{aligned}$$

$$\begin{aligned} V[X = c] &= V[X] \\ V[aX] &= a^2 V[X] \end{aligned}$$

$$\sigma_X = \sqrt{V[X]}$$

Conditional Variance

$$\begin{aligned} V[Y|X=x] &= E[(Y - E[Y|X=x])^2 | X=x] \quad \forall x \in Supp[X] \\ &= E[Y^2|X=x] - E[Y|X=x]^2 \quad \forall x \in Supp[X] \end{aligned}$$

Covariance

$$\begin{aligned} &\text{covariance} \\ Cov[X,Y] &= E[(X-E[X])(Y-E[Y])] \\ &= E[XY]-E[X]E[Y] \end{aligned}$$

If X and Y are indepdendent

$$Cov[X,Y]=0$$

If A and B are indepdendent

$$\begin{aligned} P(A\cap B) &= P(A)P(B) \\ P(A|B) &= P(A) \end{aligned}$$

Correlation

$$\rho=\frac{Cov[X,Y]}{\sigma_X\sigma_Y}$$

Sample Correlation (need to confirm the formula)

$$\hat{\rho}(X,Y)=\frac{1}{N-1}\frac{\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{S_XS_Y}$$

Estimator

Mean Squared error (MSE)

defined using a constant

$$E[(X-c)^2]=V[X]+(E[X]-c)^2$$

defined using a function

$$\begin{aligned} E[(Y-g(X))^2] &= E[E[(Y-g(X))^2|X]] \\ &= E[E[Y^2-2Yg(X)+g^2(X)|X]] \\ &= E[E[Y^2|X]-E[2Yg(X)|X]+E[g^2(X)|X]] \\ &\text{condition on X, g(X) can be treated as constant} \\ &\text{complete the square} \\ &= E[\big(E[Y^2|X]-E^2[Y|X]\big)+E^2[Y|X]-E[2Yg(X)|X]+g^2(X) \\ &= E\big[\big(E[Y^2|X]-E^2[Y|X]\big)+(E^2[Y|X]-2g(X)E[Y|X]+g^2(X))\big] \\ &\text{using the def of variance, the first term can be simplied} \\ &= E\big[V[Y|X]\big]+(E[Y|X]-g(X))^2 \\ &\text{If you are choosing some function g, you can't do better than: } E[Y|X=x] \end{aligned}$$

Break down MSE (see page 104)

$$
\begin{aligned}
&\text{MSE of Estimator} \\
E[(\theta-\hat{\theta})^2] &= V[\hat{\theta}] + \left(E[\hat{\theta}]-\theta\right)^2 \\
&= \text{variance of statistics (sampling variance)} \\
&+ \text{systematic difference between the expected value of the estimator and true value of the parameter}
\end{aligned}
$$

Expected value

$$E[x] = \sum_x x f(x)$$

Mean

- Estimator that estimates population mean, $E[X]$ based on sample
- This is a statistic.
- Differnet sample, differnet sample mean

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

Central Limit Theorem

$$
\begin{aligned}
&lim_{n\to\infty} P(\frac{Y_1+Y_2..+Y_n-n\mu}{\sigma\sqrt{n}} \le z) = \Phi(z) \\
&\text{where} \\
&Y_i \text{is iid} \\
Z &= \frac{Y_1+Y_2..+Y_n-n\mu}{\sigma\sqrt{n}} \\
&= \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}}
\end{aligned}
$$

Sample variation (z-stat)

- Variance of estimator (this case, mean)

$$
\begin{aligned}
V[\bar{X}] &= V[\frac{1}{n}(X_1 + X_2 + ... + X_n)] \\
&= \frac{1}{n^2}(V[X_1] + V[X_2] + .. + V[X_n] \\
&= \frac{V[X]}{n}
\end{aligned}
$$

where

$V[X]$ = population mean, this never changes

$V[\bar{X}]$ = Sampling variance of sample mean, this decreases as n goes up

Standard error

Estimated standard deviation of the sampling distribution

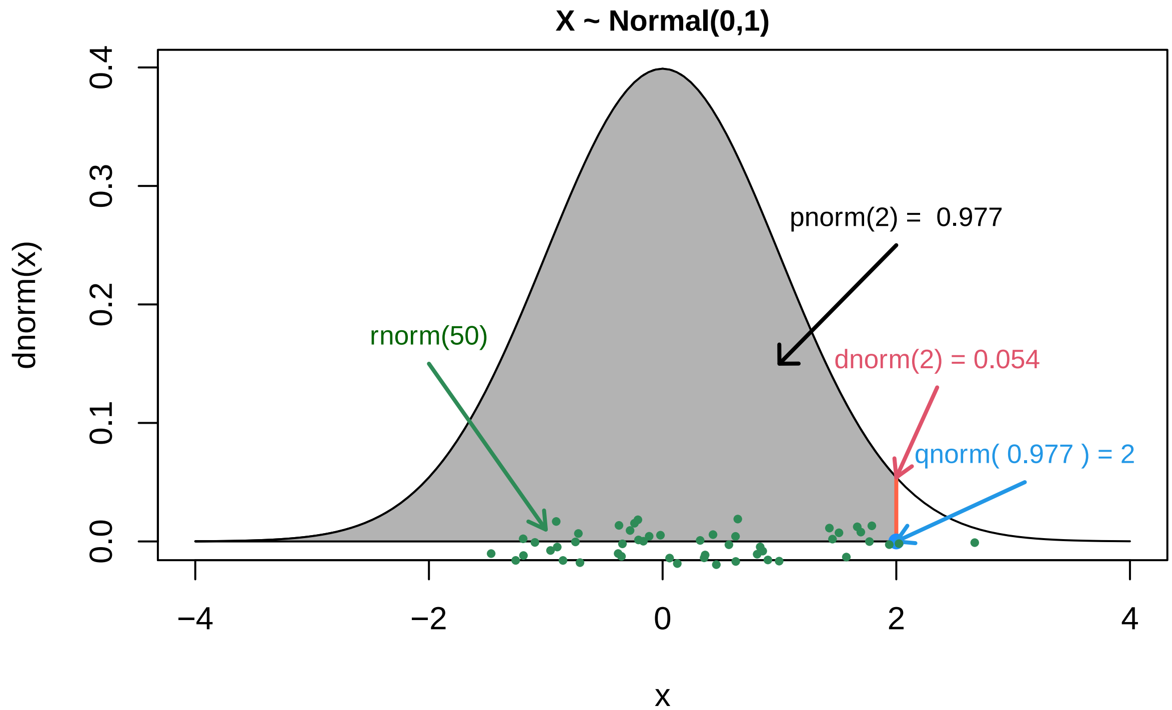
$\sqrt{Var[\bar{X}]}$

Estimated sample variation (t-stat)

- Estimated variance of estimator
- (a.k.a) estimated standard deviation of the sampling distribution

$$\hat{V}(\bar{X}) = \frac{\hat{V}(X)}{n}$$

P-value, size of test, power of test



- `p-value` is something that you observe, `size of test` is what you select and `power of test` is evaluated against another value of your parameter.
- Average human body temperature is `98.6` and body temperature is a good indication of person's health and detecting `aliens`
- suppose we know population `variance`. Then, the sample we measure can be normalized to form `z-statistic`
- Suppose we send a person whose body temperature deviates a lot from `98.6`. This is indication that the person is either `sick` or not a `human`.
  - α is `the size of the test`
  - False positive is sending someone who is healthy to hospital.
  - False negative is not sending a sick person to hospital.
  - If I send someone to hospital based on this rule, I will be sending healthy person to hospital 5% of time at maximum.

	$H_0 : \mu = 0$	$H_A : \mu \neq 0$
Not reject	Correct	
Reject	Type I Error (α)	

- Now suppose that some of the subject were actaully aliens whose mean body temperature was 103 with same variance. If I apply the same rule to these aliens, what percent of time would I be able to correct detect `aliens`?
- What percent of time would I fail to detect `aliens`?

	$H_0 : \mu = 0$	$H_A : \mu \neq 0$
Not reject	Correct	Type II Error (β)
Reject	Type I Error (α)	Correct ($1 - \beta$) power of test