Basic Math Review

Greek Letters

Greek letters are frequently used to denote either specific population properties, to name specific statistical test or as mathematical operator.

Greek Letter	Phonetic	Usage
$\boldsymbol{\alpha}$	alpha	error of first type
$\boldsymbol{\beta}$	beta	error of second type / regression parameters
$\boldsymbol{\varepsilon}$	epsilon	regression population error term
$\boldsymbol{\mu}$	mu	expected population mean
$\boldsymbol{\pi}$	pi	population probability in binomial distribution
$\boldsymbol{\rho}$	rho	population correlation coefficient
$\boldsymbol{\sigma}$	sigma	population standard deviation
$\boldsymbol{\chi}$	chi	χ^2-test
$\boldsymbol{\theta}$	theta	generic parameter of a distribution
$\boldsymbol{\lambda}$	lambda	parameter of the Poisson and exponential distributions
$\boldsymbol{\Pi}$	capital pi	multiplication symbol
$\boldsymbol{\Sigma}$	capital sigma	summation symbol

STANDARD SYMBOLS AND DEFINITION

Operation	Meaning
$\frac{Numerator}{Denominator}$	ratio between the numerator and the denominator
$\times$ *or* $\cdot$, *and* $\div$ *or* $/, +, -$	multiplication and division take precedence over addition and subtraction
$X < Y$	X is less than Y
$X \le Y$	X is less or equal than Y
$X \pm Y$	X plus minus Y, i.e., the two values $X + Y$ and $X - Y$
$\lvert X \rvert$	$X = \begin{cases} X & \text{for } X \ge 0 \\ -X & \text{for } X < 0 \end{cases}$
$\frac{1}{X} = X^{-1}$	Reciprocal of X
X^n	X to the power of n
$\sqrt{X} = X^{\frac{1}{2}}$	square root of X
$i \in \{1, 2, \dots n\}$	i is an element in the set $\{1, 2, \dots n\}$. It takes the values $1, 2,$ to n.

NOTATION FOR RANDOM VARIABLES

A random variable is denoted by a capital letter X while a lower case letter x is used to denote its observed value.

A random variable can comprise of more than values X_i relates to a specific observation. The index i ranges from $1, 2, \dots, n$. The number of observations in a variable is n. Therefore, $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$. For example, if X has $n = 4$ observation then $X = \begin{pmatrix} x_1 = 3 \\ x_2 = 5 \\ x_3 = 5 \\ x_4 = 4 \end{pmatrix}$.

RANKED DATA

- Statisticians frequently work with an ascending sorted sequence of observations which is denoted by square brackets $X_{[ranked]} = \begin{pmatrix} X_{[1]} \\ X_{[2]} \\ \vdots \\ X_{[n]} \end{pmatrix}$. For example, $X_{[ranked]} = \begin{pmatrix} x_{[1]} = 3 \\ x_{[2]} = 4 \\ x_{[3]} = 5 \\ x_{[4]} = 5 \end{pmatrix}$. Should two observations have the same rank, such as $x_i = 5$ and $x_j = 5$, then the ranks $[r]$ and $[r+1]$ will be assigned arbitrarily. See the example below:

- Ordering vectors in R:

```
> x <- c(4,5,2,3,4)
> Idx <- order(x)
> xSort <- x[Idx]
>
> cbind(x, Idx, xSort)
     x Idx xSort
[1,] 4   3     2
[2,] 5   4     3
[3,] 2   1     4
[4,] 3   5     4
[5,] 4   2     5
```

Basic Summation $\sum$ –Rules:

- $\sum_{i=1}^{n} x_i \equiv x_1 + x_2 + \cdots + x_n$. The lower index $i=1$ express the starting value of the summation sequence and the upper index n the value where the summation index i stops.
- more specifically $\sum_{i=2}^{5} x_i = x_2 + x_3 + x_4 + x_5$
- for a sum over a constant c we get $\sum_{i=1}^{n} c = n \cdot c$

- for a mixture of a constant and a variable $\sum_{i=1}^{n} c \cdot x_i = c \cdot \sum_{i=1}^{n} x_i$
- for an additive mixture of variables $\sum_{i=1}^{n} \left(x_i + y_i \right) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
- Inequalities (so do not confuse either side of the expression; they lead to different results):

  $\sum_{i=1}^{n} x_i \cdot y_i \neq \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i$

  $\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i \right)^2$

- Special rule for ranks: $\sum_{i=1}^{n} i = \frac{n}{2} \cdot (n+1)$
- Doubly index variables x_{ij} in a cross-tabulation (or matrix) with I rows and J columns:

  Let: $\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1J} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{iJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ij} & \cdots & x_{IJ} \end{bmatrix}$

  Then the i^{th} row sum is $x_{i+} = \sum_{j=1}^{J} x_{ij}$ and the j^{th} column sum is $x_{+j} = \sum_{i=1}^{I} x_{ij}$ and the total sum becomes

  $x_{++} = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij} = \sum_{i=1}^{I} x_{i+} \text{ or } \sum_{j=1}^{J} x_{+j}$

- The R functions:
  - **`sum()`** calculated the sum over the elements of a vector
  - **`rowSums()`** calculates along the rows of a matrix a vector of row sums.
  - **`colSums()`** calculates along the columns of a matrix a vector of column sums.
- <u>Example:</u> The variance estimator can either be calculated by $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$ or by $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} \cdot \bar{x}^2$. To derive this equivalence of both expressions, remember the definition of the mean $\bar{x} = 1/n \cdot \sum_{i=1}^{n} x_i$:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i^2 - 2x_i \cdot \bar{x} + \bar{x}^2)$$

$$= \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2 \cdot x_i \cdot \bar{x} + \underbrace{\sum_{i=1}^{n} \bar{x}^2}_{=n\cdot\bar{x}^2} \right) = \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_i^2 - 2 \cdot \bar{x} \cdot \underbrace{\sum_{i=1}^{n} x_i}_{=n\cdot\bar{x}} + n \cdot \bar{x}^2 \right)$$

$$= \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_i^2 \underbrace{-2 \cdot n \cdot \bar{x}^2 + n \cdot \bar{x}^2}_{-n\cdot\bar{x}^2} \right) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} x_i^2 - \frac{n}{n-1} \cdot \bar{x}^2$$

Finding the Minimum of a Quadratic Function

- In statistic, we encounter frequently the need to find an optimal value of a function. If we want to minimize square deviations around an unknown value, the optimal value would be the minimum.
- The minimum is found at that point where the slope of the function is zero. The slope of a function is measured by the first derivative.
- Basic rules of derivatives:

$$\frac{\partial}{\partial x} f(a) = 0 \quad \text{The function } f(a) \text{ is constant with regards to } x$$

$$\frac{\partial}{\partial x} a \cdot x^n = a \cdot n \cdot x^{n-1} \quad \text{Example: } \frac{\partial}{\partial x} 3 \cdot x^2 = 6 \cdot x$$

$$\frac{\partial}{\partial x}\left(f(x) + g(x) \right) = \frac{\partial}{\partial x} f(x) + \frac{\partial}{\partial x} g(x) \quad \text{Example: } \frac{\partial}{\partial x}\left(3 \cdot x^2 + 5 \cdot x^{-1}\right) = 6 \cdot x - 1 \cdot 5 \cdot x^{-2}$$

- Which value of θ (theta) minimizes the quadratic expression $\min_{\theta} \sum_{i=1}^{n} \left(x_i - \theta \right)^2$?

  $f(\theta) = \sum_{i=1}^{n} \left(x_i - \theta \right)^2 = \sum_{i=1}^{n} x_i^2 - 2 \cdot \theta \cdot \sum_{i=1}^{n} x_i + n \cdot \theta^2$

  Take the first derivative with regard to θ, which is the slope of $f(\theta)$ at θ:

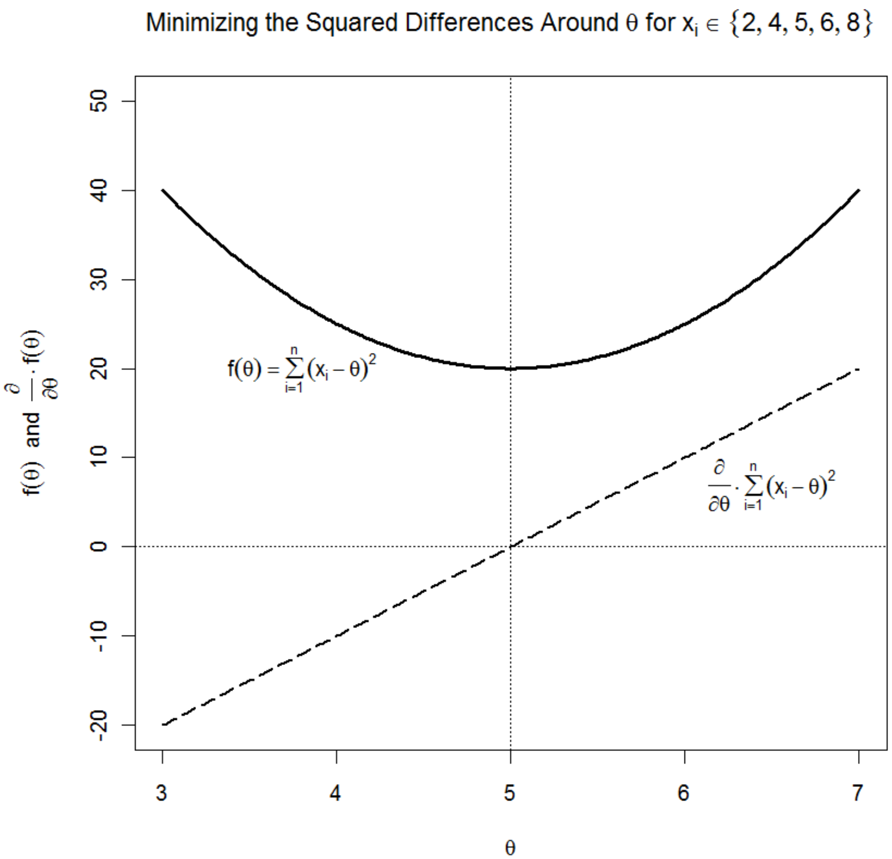
  $$\frac{\partial}{\partial \theta} \left(\underbrace{\sum_{i=1}^{n} x_i^2}_{\text{does not depend on } \theta} - 2 \cdot \theta \cdot \sum_{i=1}^{n} x_i + n \cdot \theta^2 \right) = -2 \cdot \sum_{i=1}^{n} x_i + 2 \cdot n \cdot \theta$$

  At its maximum or minimum the first derivative (that is, the slope) is zero for a given θ.

  We thus get: $-2 \cdot \sum_{i=1}^{n} x_i + 2 \cdot n \cdot \theta = 0 \Leftrightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{n}$

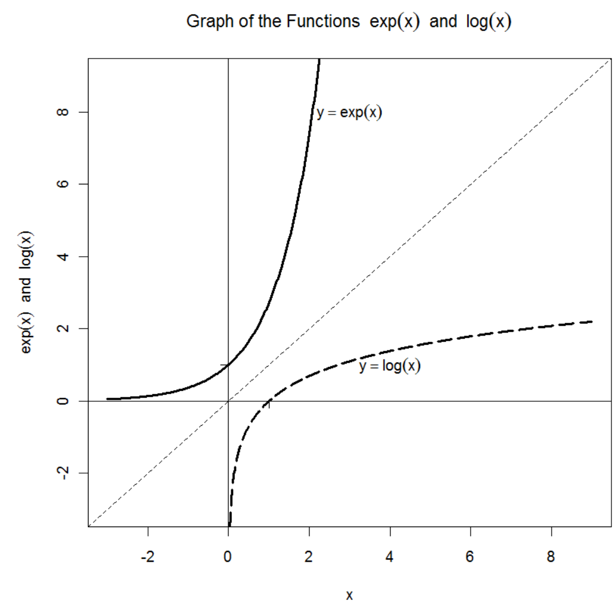
  $\Rightarrow$ This is the well-know arithmetic mean!!!

- <u>Example:</u> The data values are $x_i \in \{2, 5, 4, 6, 8\}$. Thus the function to be minimized with respect to θ is $f(\theta) = \left(2 - \theta\right)^2 + \left(5 - \theta\right)^2 + \left(4 - \theta\right)^2 + \left(6 - \theta\right)^2 + \left(8 - \theta\right)^2$



The solution is found at $\theta = 5 \Leftrightarrow \bar{x}$.

The Exponential and Logarithmic Functions



- Both functions are inversely related: $x = \exp(\log(x))$ and $x = \log(\exp(x))$.
  Notes:
  - The $\log -$function is usually the natural logarithm to the basis of the Euler constant $e = 2.718$
  - The support of the logarithmic function is limited from below by zero, that is, $x \in]0, \infty]$ with $\log(0) = -\infty$.
- Both functions ***distort*** constant distance units of the variable x.
  E.g., $\Delta x_1 = 4 - 2 = 2$ and $\Delta x_2 = 8 - 6 = 2$
  but $\log(4) - \log(2) = 1.086$ and
  $\log(8) - \log(6) = 0.288$, respectively.
  This means, logarithmic distances at the upper end of the scale become shorter.
- Basic rules:
  - Logarithmic function: $\log(x \cdot y) = \log(x) + \log(y)$, $\log(x/y) = \log(x) - \log(y)$ and $\log(x^y) = y \cdot \log(x)$
  - Exponential function: $\exp(x + y) = \exp(x) \cdot \exp(y)$, $\exp(x - y) = \exp(x)/\exp(y)$ and $[\exp(x)]^y = \exp(x \cdot y)$

Appendix 1: Population and Sampling Distributions (Ham pp 289-293)

- In ***theoretical statistics*** we make statements about the population based on the distribution $f(y)$ of a ***continuous*** random variable Y or $\Pr(Y=y)$ for a ***discrete*** random variable Y, which takes the specific value y, respectively.
- In applied statistics we are dealing with sampled data from the population and aim at estimating properties of the underlying population from which the random sample has been drawn
- The sample is our narrow keyhole allowing us to look at parts of the unknown population.
- **Conventions:**
  - Parameters characterizing the population are usually denoted by Greek characters, e.g., the expectation μ_X of the random variable *X*. Their estimates are either expressed by Latin characters, e.g., the mean $\bar{X}$, or by a hat symbol that denotes an estimate, e.g., $\hat{\mu}_X$.
  - A random variable from the population is usually denoted by a capital letter, e.g., X, whereas its observed realization in the sample is denoted by small letters, e.g., $x_1, x_2, \ldots, x_n$
- **Population expectation (central tendency)**
  - The mean in the unknown population is called ***expectation*** and denoted by $E[X]=\mu_X$
  - For ***discrete*** variables the expectation function is defined by $E[Y]=\sum_{i=1}^{I} y_i \cdot \Pr(Y=y_i)$

where I is the total number of representations, which can be an infinite number as for the Poisson distribution $y_i \in \{0,1,2,\ldots,\infty\}$ or a finite set as in the sum of two throws of a dices $y_i \in \{2,3,\ldots,12\}$

- For ***continuous*** random variables the expectation function is defined in terms of the density function $f(x)$

  $$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$$

  For infeasible values of x the density will become $f(x) = 0$ because these values are improbable.

- Remember: the density function $f(x)$ at x cannot be interpreted as probability. We only can express the probability for a range of value:

  $X \in [a, b] \Rightarrow \Pr(a \leq X \leq b) = \int_a^b f(x) \cdot dx.$

- Some rules for the expectation:

  $E[a] = a$ for a deterministic (constant) value a

  $$E[a \cdot X] = a \cdot E[X]$$

  $$E[X \pm Y] = E[X] \pm E[Y]$$

  $$E[a + b \cdot X] = a + b \cdot E[X]$$

  $$\begin{aligned} E[a \cdot X + b \cdot Y] &= E[a \cdot X] + E[b \cdot Y] \\ &= a \cdot E[X] + b \cdot E[Y] \end{aligned}$$

- An unbiased sample estimator of the expectation $E[Y]$ is the mean

  $$\bar{Y} = \tfrac{1}{n} \cdot \sum_{i=1}^{n} y_i$$

- **Variance**
  - The variance is a measure of squared spread around the center, i.e., expectation, of a random variable
    $$Var[X] = \int \left(x - E[X]\right)^2 \cdot f(x) \cdot dx$$
    $$= E\left[\left(X - E[X]\right)^2\right]$$
    $$= E\left[X^2\right] - \left(E[X]\right)^2$$
  - The unbiased sample variance estimator s_X^2 for the population variance σ^2 is:
    $$s_X^2 = \tfrac{1}{n-1} \cdot \sum_{i=1}^{n} \left(x_i - \bar{X}\right)^2$$
  - Basic properties:
    $Var[a] = 0$ because a is a constant (i.e., not random)
    $$Var[b \cdot X] = b^2 \cdot Var[X]$$
    $$Var[X + Y] = Var[X] + Var[Y] + 2 \cdot Cov[X, Y]$$
    $$Var[X - Y] = Var[X] + Var[Y] - 2 \cdot Cov[X, Y]$$
    $$Var[a + b \cdot X] = Var[a] + Var[b \cdot X]$$
    $$= b^2 \cdot Var[X]$$
    $$Var[a \cdot X + b \cdot Y] = a^2 \cdot Var[X] + b^2 \cdot Var[Y] + 2 \cdot a \cdot b \cdot Cov[X, Y]$$

Example: Explanation of Integration using The Exponential Distribution

- Background information on the exponential distribution
  - Example: the ***waiting times*** x between two independent random events (earth quakes, customers lining up in-front of a cashier etc.) may be exponential distributed.
  - The exponential distribution only has the one parameter λ, with $E[X] = 1/\lambda$ being the average waiting time.
  - You can look at some exponential distributions using **`dexp( )`** function.
  - The exponential distribution is related to the ***Poisson*** distribution:
    - It provides a stochastic model for the number of independent random events y within a fixed time-interval.
    - The expected number of random events within a fixed time-interval is $E[Y] = \lambda$.
    - If the expected number of events is large the average waiting time between the events will be small.
      Thus we have an inverse relationship between both expectations.
  - The density function of the exponential distribution is

$$f(x \mid \lambda) = \begin{cases} \lambda \cdot \exp(-\lambda \cdot x) & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

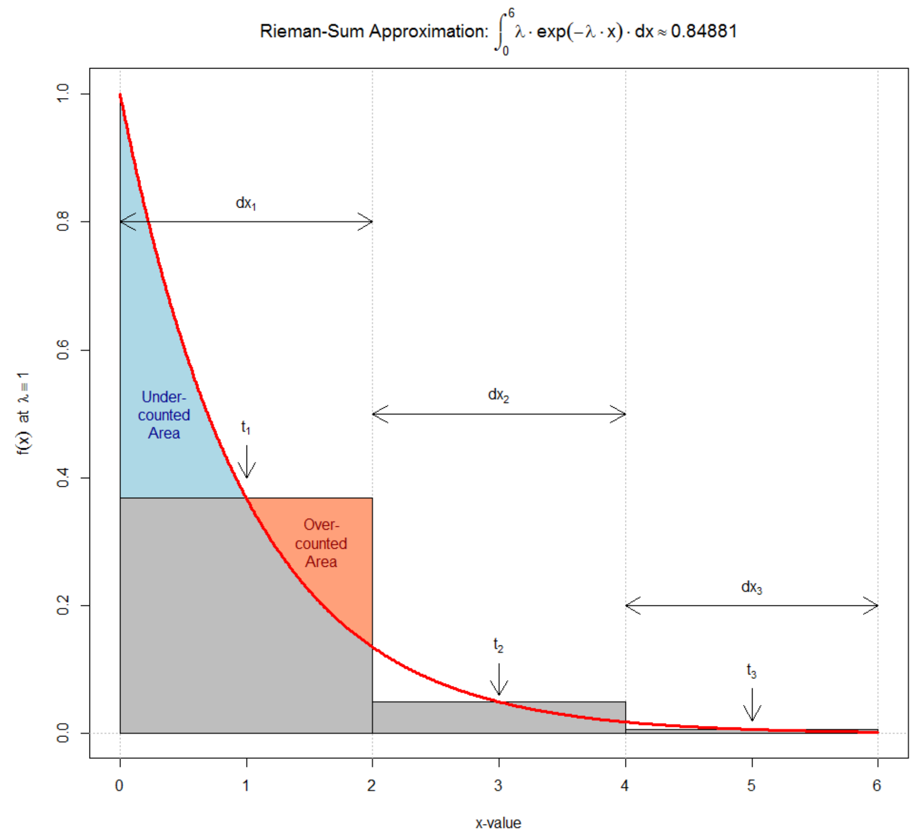
- Its cumulative distribution function is

$$F(x\,|\,\lambda)=\int_0^x \lambda\cdot\exp(-\lambda\cdot x)\cdot dx=\begin{cases}1-\exp(-\lambda\cdot x) & \text{for } x\geq 0\\ 0 & \text{otherwise}\end{cases}$$

- Its moments are known analytical:
  - The expectation is $E[X]=\int_0^\infty x\cdot\lambda\cdot\exp(-\lambda\cdot x)\cdot dx=\dfrac{1}{\lambda}$ and
  - the variance is $Var[X]=\int_0^\infty\left(x-\underbrace{1/\lambda}_{=E[X]}\right)^2\cdot\lambda\cdot\exp(-\lambda\cdot x)\cdot dx=\dfrac{1}{\lambda^2}$
- The parameter λ can be estimated from sample observations by $\hat{\lambda}=1/\bar{x}$.

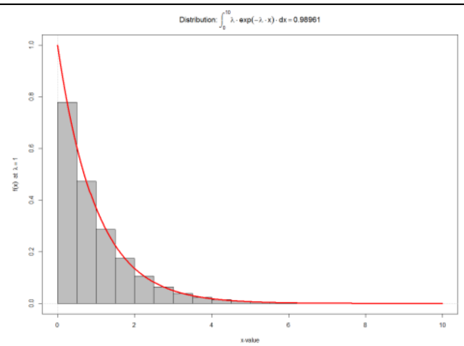
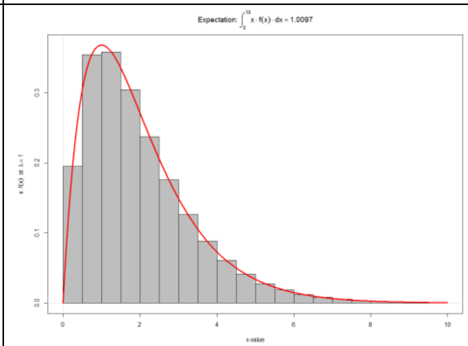
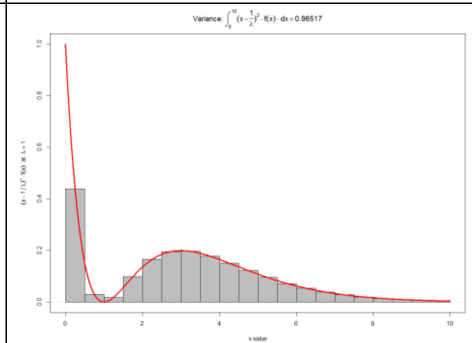
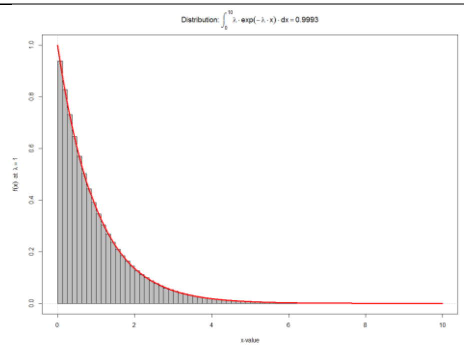
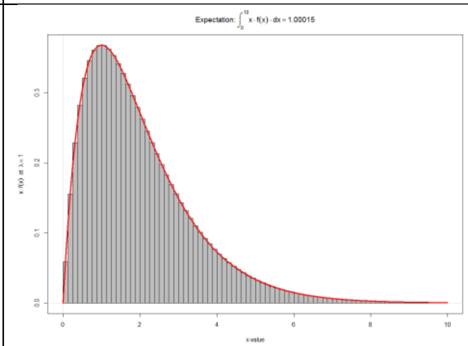
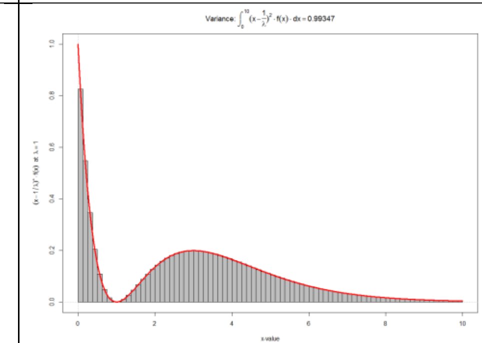
- Evaluation of the moments by numerical integration (see script **RIEMANNSUM.R**):
  - The Riemann sum approximates a continuous integral by $\int_a^b f(x)\cdot dx\approx\sum_{i=1}^n f(t_i)\cdot dx_i$ by discrete evaluations with $a=x_0<x_1<x_2<\cdots<x_{n-1}<x_n=b$, with the bin width $dx_i=x_i-x_{i-1}$ and $t_i\in[x_{i-1},x_i]$, which usually is set to the halfway point $t_i=\dfrac{x_{i-1}+x_i}{2}$

- The parameters dx_i and n determine the resolution and therefore the accuracy of the Riemann sum integral approximation.
- Advance integration algorithms make the differences $dx_i = x_i - x_{i-1}$ adaptive relative to the variability of $f(x)$:
  - If the underlying function $f(x)$ varies heavily, then the differences dx_i should be small.
  - On the other hand, if the underlying function is fairly smooth the differences dx_i could larger.



The underlying idea is similar to an adaptive kernel density estimator.

- Evaluation of the exponential density, expectation and variance for $\lambda = 1$ in the range $x \in [0,10]$:

Number of bins n	$f(x)$ (density)	$x \cdot f(x)$ (expectation)	$(x - E[X])^2 \cdot f(x)$ (variance)
$n = 20$			
$n = 80$			

- Notes:
  - The integral $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$ over any ***density functions*** $f(x)$ always is one.

- ***Theoretically*** all integrals in the example should be equal to one, because $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$, $E(X) = 1/\lambda$ and $Var(X) = 1/\lambda^2$ for an exponential distribution with $\lambda = 1$.
- Even if we increase the number of bins, these integrals will ***not*** approach 1 because we are ***truncating*** infinitive integration range by the upper value $b = 10 < \infty$.

- **Covariance**
  - The covariance is a basic measure of the ***linear relationship*** between pairs of random variables.
  - The covariance is the numerator of the correlation coefficient. That is $\rho = \frac{Cov[X,Y]}{\sqrt{Var[X] \cdot Var[Y]}}$
  - The covariance of a variable with itself is called the variance: $Cov[X,X] = Var[X]$
  - $$\begin{aligned} Cov[X,Y] &= \iint \left(x - E[X]\right) \cdot \left(y - E[Y]\right) \cdot f(x,y) \cdot dx \cdot dy \\ &= E\left[\left(X - E[X]\right) \cdot \left(Y - E[Y]\right)\right] \\ &= E[X \cdot Y] - E[X] \cdot E[Y] \end{aligned}$$
  - An unbiased estimator for the population covariance is $s_{XY} = \frac{\sum_{i=1}^{n} \left[\left(x_i - \bar{X}\right) \cdot \left(y_i - \bar{Y}\right)\right]}{n-1}$
  - Some rules:
    $Cov[a,Y] = 0$
    $Cov[b \cdot X, Y] = b \cdot Cov[X,Y]$
    $Cov[X + W, Y] = Cov[X,Y] + Cov[W,Y]$

- The covariance is unaffected by the addition of a constant to either random variable:
  $$Cov[a+X,Y] = \underbrace{Cov[a,Y]}_{=0} + Cov[X,Y]$$
  $$= Cov[X,Y]$$
- The covariance between sums of variables reduces to sums of covariances between their components
  $$Cov[X+W,Y+Z] = Cov[X+W,Y] + Cov[X+W,Z]$$
  $$= Cov[X,Y] + Cov[W,Y] + Cov[X,Z] + Cov[W,Z]$$
  $$Cov[X,Y-X] = Cov[X,Y] - Cov[X,X]$$
  $$= Cov[X,Y] - Var[X]$$

- The Ordinary Least Squares slope estimator in terms of covariances becomes
  - The slope regression coefficient for a regression of Y onto X becomes
    $$\beta_{1,Y|X} = \frac{Cov[X,Y]}{Var[X]}$$
  - Vice versa, for a regression of X onto Y one gets $\beta_{1,X|Y} = \dfrac{Cov[X,Y]}{Var[Y]}$
  - The regression intercept for a regression of Y onto X becomes $\beta_{0,Y|X} = E[Y] - \beta_{1,Y|X} \cdot E[X]$, because the expectations $E[Y]$ and $E[X]$ lie on the regression line.

Normal Distribution and its Relatives

- Definition: Let z and the sets $z_1, z_2, \ldots, z_n$ with n elements and $\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_m$ with m elements be ***standard normal*** distributed random variables which are all ***mutually independent***.
- The χ^2-distribution: The random variables
$$s_n^2 = \sum_{i=1}^{n} z_i^2 \text{ and } \tilde{s}_m^2 = \sum_{i=1}^{m} \tilde{z}_i^2$$
of the ***sums of squared*** independent standard normal distributed variables are χ^2-distributed
$$s_n^2 \sim \chi^2_{df=n} \text{ and } \tilde{s}_m^2 \sim \chi^2_{df=m}$$
with n and m degrees of freedom, respectively.
The expected value of a χ^2-distributed variable is equal to its degrees of freedom.
- The t-distribution: Let $t_n = \frac{z}{\sqrt{s_n^2/n}}$ and $\tilde{t}_m = \frac{z}{\sqrt{\tilde{s}_m^2/m}}$ with z being independent standard normal distributed. Then t_n and $\tilde{t}_m$ are t-distributed with with n and m degrees of freedom, respectively.
- The F-distribution: Let $F_n^m = \frac{s_n^2/n}{\tilde{s}_m^2/m}$. Then F_n^m is F-distributed with n and m degrees of freedom.

Bias and Mean Square Error

- The theoretical sampling distribution of a statistic $\hat{\theta}$ is evaluated over all possible random samples of a given size n.

- A statistic is ***unbiased*** if $E[\hat{\theta}] = \theta$, that is, $E[\hat{\theta}] - \theta = 0$. It is biased if $E[\hat{\theta}] \neq \theta$, that is, the estimator's $\hat{\theta}$ expected value differs from the true population parameter θ.
- The variance $Var[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2]$ expresses the precision of a sample statistics. The square root of this variance is called ***standard error*** of a sample statistics.
- The mean square error is

$$\begin{aligned} MSE &= E\left[(\hat{\theta} - \theta)^2\right] \\ &= E\left[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta)^2\right] \\ &= E\left[(\hat{\theta} - E[\hat{\theta}])^2 + 2 \cdot (\hat{\theta} - E[\hat{\theta}]) \cdot (E[\hat{\theta}] - \theta) + (E[\hat{\theta}] - \theta)^2\right] \\ &= Var[\hat{\theta}] + bias^2 \end{aligned}$$

with the term $E[2 \cdot (\hat{\theta} - E[\hat{\theta}]) \cdot (E[\hat{\theta}] - \theta)] = 0$ because

$$\begin{aligned} E[(\hat{\theta} - E[\hat{\theta}]) \cdot (E[\hat{\theta}] - \theta)] &= E[\hat{\theta} \cdot E[\hat{\theta}] - \hat{\theta} \cdot \theta - E[\hat{\theta}] \cdot E[\hat{\theta}] + E[\hat{\theta}] \cdot \theta] \\ &= E[\hat{\theta}] \cdot E[\hat{\theta}] - E[\hat{\theta}] \cdot \theta - E[\hat{\theta}] \cdot E[\hat{\theta}] + E[\hat{\theta}] \cdot \theta \\ &= 0 \end{aligned}$$

Only $\hat{\theta}$ is a random variable, whereas $E[\hat{\theta}]$ and θ are constants and therefore, $E\left[E[\hat{\theta}]\right] = E[\hat{\theta}]$ and $E[\theta] = \theta$.